7-11-2013

Improvements on Manual Pavement Distress Data Collection to Conform to State and Federal Requirements

Kelly Montoya

Kelly Montoya

Candidate

Civil Engineering

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Dr. Susan Bogus Halter , Chairperson

Dr. Vanessa Valentin

Dr. Paola Bandini

Improvements on Manual Pavement Distress Data Collection to Conform to State and Federal Requirements

by

Kelly Montoya

BS, Civil Engineering, University of New Mexico, 2010

THESIS

Submitted in Partial Fulfillment of the Requirements

for the Degree of

Masters of Science,

Civil Engineering

University Of New Mexico

Albuquerque, New Mexico

May 2013

Acknowledgements

I could not have done this without support from so many people who helped me and made sure a project of this size could be done in only one year.

Dr. Giovanni Migliaccio, who encouraged me that I could get into the Master's Program and work on this project.

My wonderful project coordinators, Dr. Susan Bogus Halter and Dr. Paola Bandini, who had the experience and knowledge necessary to guide me through the process and make sure I was successful. I can't find words to express my gratitude.

The New Mexico Department of Transportation who sponsored the project, NM10MNT-01(Bandini et al 2012), including the Research Bureau in Albuquerque and District 1 Maintenance crews who made it safe for the students during road closures to collect the necessary data.

All of the students who worked tirelessly to collect statewide data and help out with data processing:

- Hung V. Pham of NMSU – his talent with data processing is unmatched

- Jacob Zagone, Justin Lindsey, and Joshua J. Alonzo of NMSU, who collected data in the south side of the state

- Seth Lewis of UNM who collected data in the north side of the state.

My very special thanks to David Barboza of UNM, who did what I told him to do for a whole year, I really couldn't have done it alone.

Improvements on Manual Pavement Distress Data Collection to Conform to State and Federal Requirements

by

Kelly Montoya

B.S., Civil Engineering, University of New Mexico, 2010

M.S., Civil Engineering, University of New Mexico, 2013

Abstract

The purpose of this study is to align the pavement distress data of the state of New Mexico's pavement management system with that of the federal highway system, the Highway Performance Monitoring System (HPMS), while concurrently reducing variability, increasing data consistency, and providing a greater quantity of specific data items to be considered in pavement preservation and keeping costs under control. The distress rating codes used by NMDOT in reporting severity and extent do not readily correlate to the data required in the HPMS system. To form a usable data set for both NMDOT's PMS and the federal HPMS, an entirely new distress data collection protocol has been written and field tested. The three levels of testing include interrater agreement and interrater reliability across the "old" protocol, the "new" protocol, and actual field measurements to judge the accuracy of both old and new methods. By doing this, the accuracy and validity of the pavement distress surveys are expected to increase. The objectivity and integrity of the NMDOT distress rating criteria for flexible pavements should also be increased through a revision of the old criteria.

Table of Contents

Chapter 1 Introduction 1

1.1 Overview 1

1.2 Background 2

1.2.1 Pavement Preservation 2

1.2.2 The Highway Performance Monitoring System (HPMS) 4

1.2.3 The New Mexico Department of Transportation (NMDOT) Pavement

1.3 Research Objective 7

Chapter 2 Background 9

2.1 Pavement Management 9

2.1.1 Data Collection 9

2.1.2 Distress Types 12

2.1.3 Administrative Levels of Pavement Management 15

2.1.4 Users of Pavement Management Systems 17

2.1.5 Maintenance/Repair/Reconstruction 18

2.1.5 Pavement Condition Indexes 21

2.2 Future Directions of Pavement Management Systems 22

2.3 Statistical Analysis of Data Variability 23

2.3.1 Fully Automated Data Collection 24

2.3.2 Software Improvements 25

2.3.3 Geographic Information Systems 25

Chapter 3 Methodology 27

3.1 Old Data Collection Protocol 27

3.2 Development of New Data Collection Protocol 30
3.3 Testing of Old versus New Data Collection Protocols 38
3.3.1 Initial Data Collection 38
3.3.2 Interrater Agreement 42
3.3.3 Detailed Field Measurements 47
Chapter 4 Results and Discussion 57
4.1 Interrater Agreement Results 57
4.1.1 Interrater Agreement - Old Method 57
4.1.2 Interrater Agreement – New Method 59
4.1.3Interrater Agreement Comparison Between Protocols 76
4.2 Agreement Over Time 77
4.3 Rating Estimation vs. Actual Measurements 88
4.3.1 Transverse Cracking Analysis 88
4.3.2 Effect of Crack Geometry 92
4.3.3 Alligator Cracking Analysis 94
4.4 Distress Rate Analysis 97
4.5 Time Required for Visual Surveys 102
Chapter 5 - Conclusions 105
5.1 Conclusions 105
5.1.1 Observations 106
5.1.2 Time-History Data Collection Issues 106
5.1.3 Pavement Condition Data Consistency 107
5.2 Implementation 108

5.3 Future Research 113
Appendices... 115

Appendix A – Old Field Form 116

Appendix B – Old Rating Criteria 117

Appendix C – New Field Form 118

Appendix D – New Rating Criteria 119

List of Tables

Table 1: Number of Respondents using Data Collection Methods 10

Table 2: Weighting Factors for Flexible Pavement Distresses 30

Table 3: List of Mileposts 40

Table 4: Nevada Transverse Cracking Limits 46

Table 5: Test Sections for Field Measurements of Fatigue and Transverse Cracking 50

Table 6: Alligator Cracking Extent Conversion 64

Table 7: Sample Data for Alligator Cracking 65

Table 8: Extent Ranges for Transverse Cracking 72

Table 9: Sample Data for Transverse Cracking 72

Table 10: Field Measurements and Estimates of Transverse Cracking Length 89

Table 11: Transverse Cracking 2-Point and Actual Length Field Measurements 92

Table 12: Field Measurements for Alligator Cracking Analysis 95

Table 13: Rater 1 – Evaluation Times 103

List of Figures

Figure 1: Pavement Preservation Philosophy 3

Figure 2: Top-Down Schematic 16

Figure 3: Bottom-Up Schematic 16

Figure 4: Types of Pavement Preservation 19

Figure 5: Average Extent of Raveling & Weathering 35

Figure 6: Average Extent of Bleeding 36

Figure 7: Chalk Markings in Preparation for Measurements 52

Figure 8: Actual vs. Straight-line Length of Transverse Cracks 53

Figure 9: Outlining Cracking Area 55

Figure 10: Grid Application 55

Figure 11: Alligator Cracking Instance 56

Figure 12: Measuring Alligator Cracking Using Grid 56

Figure 13: Summary of Round 1 58

Figure 14: Summary of Round 2 58

Figure 15: Alligator Cracking AD(M) by Severity and Round 61

Figure 16: Alligator Cracking AD(Md) by Severity and Round 61

Figure 17: Total Alligator Cracking, Round 1 62

Figure 18: Total Alligator Cracking, Round 2 63

Figure 19: Translated AD(M) and AD(Md) for Alligator Cracking 66

Figure 20: Figure 20: Total Alligator Cracking by Rater 67

Figure 21: Transverse Cracking AD(M), Through 2 Rounds 68

Figure 22: Transverse Cracking AD(Md), Through 2 Rounds 69

Figure 23: Total Transverse Cracking Round 1 69

Figure 24: Total Transverse Cracking Round 2 70

Figure 25: Translated Transverse Cracking AD(M) and AD(Md) By Round 73

Figure 26: Raveling & Weathering AD(M) and AD(Md) 74

Figure 27: Bleeding AD(M) and AD(Md) By Round 75

Figure 28: Longitudinal Cracking, Rounds 1 and 2 76

Figure 29: Averaged AD(Md) – Old and New Protocols 77

Figure 30: Total Alligator Cracking by Rater and Severity 78

Figure 31: Agreement over Time - Alligator Cracking 79

Figure 32: Transverse Cracking by Rater and Severity 80

Figure 33: Total Transverse Cracking by Rater 81

Figure 34: Agreement over Time- Transverse Cracking 82

Figure 35: Total Raveling and Weathering by Rater 83

Figure 36: Agreement over Time – Raveling and Weathering 83

Figure 37: Total Bleeding Summary by Rater 85

Figure 38: Agreement over Time – Bleeding 85

Figure 39: Total Longitudinal Cracking by Rater 86

Figure 40: Agreement over Time – Longitudinal Cracking 87

Figure 41: Raters' Estimated vs. Actual Length, All Transverse Cracking 91

Figure 42: Raters' Estimated vs. Actual Length, Cracks > 6 ft 91

Figure 43: Transverse Cracking Measurements: Actual vs. 2-point Lengths 94

Figure 44: Alligator Cracking, Esimated vs. Detailed Measurements 97

Figure 45: Distress Rate Comparison – Old Method 99

Figure 46: Distress Rate Comparison – Translated Protocol 100

Figure 47: Distress Rate Comparison – New vs. Old, Round 1 101

Figure 48: Distress Rate Comparison – New vs. Old, Round 2 101

Chapter 1

Introduction

This research aims to improve upon an existing method of manual data collection that the New Mexico Department of Transportation (NMDOT) uses in order to assess the condition of their pavement network. Because federal reporting requirements changed in 2010, it was important to make adjustments in the way data are collected in order to simplify the process and provide a data set that can be used to fulfill both state and federal mandates. A new protocol for this collection was written, tested for consistency, and validated in the field using actual pavement sections to calibrate the results and verify the accuracy of the new protocol.

1.1 Overview

Since the advent of the automobile in the early 20th century, there has been a demand for high quality roads. By the 1950's, the construction of the Interstate Highway System was well on its way, and road users began demanding more roads at even higher quality. Many of the first roads built are obsolete by today's standards, as pavements built today can be engineered to meet a variety of needs—greater durability, enhanced skid resistance to improve safety; and, a smoother ride to satisfy the traveling public.

According to the National Center for Pavement Preservation (2010), there are 2,319,535 miles of paved roads in the United States representing billions of dollars in investments. Now, we are faced with a huge problem. The vast system of pavement that has been built over the last 50 years is in various states of disrepair, and the current economic state is not conducive to nationwide maintenance and repairs. This is the reason state and federal agencies rely on pavement management systems to make informed decisions on which road sections should receive priority attention.

The actual concept of pavement management has been around for some time. Managerial decisions are made every day as part of normal daily operating measures. In the past, however, pavements were maintained, but not truly managed. It was determined by a local pavement engineer which roads should be repaired or maintained and on what schedule, regardless of any other pavement condition nearby or with any regard to life cycle cost analysis. Pavement management seeks to improve the efficiency of decision making regarding pavement design, maintenance, and repair, expanding its scope, ensuring the ramifications of all decisions are known, and increasing consistency (Haas, 1987). Technological advances have given agencies the tools by which to manage pavements economically. A true Pavement Management System (PMS) provides a systematic, consistent method for selecting maintenance and repair needs and determining priorities and the optimal time of repair by predicting future pavement condition (Shahin, 1994).

1.2 Background

1.2.1 Pavement Preservation

Perhaps the fastest growing area of pavement management is pavement preservation. Pavement preservation is a proactive approach in maintaining highways. Rather than waiting and allowing a road to require reconstruction after (or even before) its design life, a series of activities can be done at scheduled intervals (or condition trigger points) while the pavement is in good condition to extend remaining service life beyond the design period (FHWA, 2011). The most commonly-used diagram to describe the pavement preservation philosophy is shown in Figure 1 (derived from similar diagrams from AASHTO and FHWA).

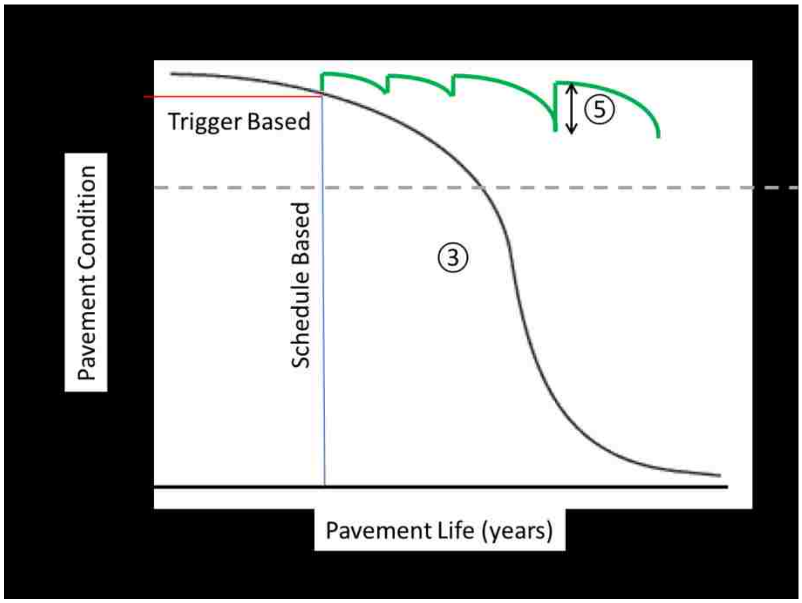


Figure 1: Pavement Preservation Philosophy (adapted from AASHTO and FHWA)

The areas represented on the diagram can be described as follows:

1. The vertical axis is based on the condition of the pavement. This can be defined by several indices, such as the Pavement Condition Index, Pavement Serviceability Index, distress rate, or any other index that gives a numerical value based on age and distress or agency-specific factors.
2. The horizontal axis is based on time, mostly pavement life, but it can be defined in terms of remaining service life.
3. This line is the pavement performance curve. These curves are agency generated and based on construction standards, mix designs, environmental factors, and other similar factors affecting performance. Most pavements behave in the

fashion shown in Figure 1 – when they begin to deteriorate, they deteriorate fast. The gray dashed line represents a threshold that denotes a change in cost to bring pavements up to an acceptable condition. Generally, repairs done above the line cost from 1/4 to 1/10 of those performed after the pavement has degraded beyond the threshold.

4. This part of the graph shows the extension of the life of a pavement when treatment is performed. In general, the longer the line, the more usable life is gained.
5. This vertical line represents the gain in pavement condition as a result of the treatment. Here, a longer line signifies a greater increase in pavement condition.

The red line indicates a condition parameter that will trigger some sort of preventive maintenance. When the observed distresses cause a drop in the index used and that drop reaches a particular value, maintenance is required. The blue line represents schedule-based repairs. This parameter simply indicates that after a pavement reaches a certain age, regardless of condition, maintenance is performed. The key concept in deciding when these triggers should take place is to ensure the pavement has not deteriorated to the point where reactive maintenance is required. An example of reactive maintenance is to have to perform a mill and inlay rather than a simple fog seal because the pavement surface has declined past a borderline functional performance.

1.2.2 The Highway Performance Monitoring System (HPMS)

The HPMS is sponsored by the Federal Highway Administration (FHWA) and is a national level highway information system. It contains data on nearly all characteristics of the nation's highways: performance, use, operating qualities, extent, and condition. All public roads are included in the HPMS on the administrative and extent information database, while major arterials and collectors exist as a mix of system-wide and sample data. (FHWA, 2010) The lowest functional systems have the most limited data, as the travel is only summarized.

The major purpose of the HPMS is to support a data driven decision process within FHWA, the DOT, and the Congress. The HPMS data are used extensively in the analysis of highway system condition, performance, and investment needs that make up the biennial Condition and Performance Reports to Congress. These Reports are used by the Congress in establishing both authorization and appropriation legislation, activities that ultimately determine the scope and size of the Federal-aid Highway Program, and determine the level of Federal highway taxation.

These data are also used for assessing changes in highway system performance brought about by implementing funded highway system improvement programs. HPMS is a national, unique source of highway system information that is made available to those in the transportation community for highway and transportation planning and other purposes through the annual Highway Statistics and other public distribution media.

It is the responsibility of each state’s DOT to submit all public roadway mileage that is consistent with each state’s Certified Public Mileage to the FHWA by June 15 of each year. FHWA supplies guidelines and examples in detail for this process in their field manual. To achieve a quality program it is vital that systems be established to accurately collect and maintain internal data in accordance with these guidelines as well as establish communications and mechanisms with municipal and federal agencies to properly maintain and report their data.

1.2.3 The New Mexico Department of Transportation (NMDOT) Pavement Management System

The NMDOT has collected pavement condition data (i.e., surface distresses, rutting and roughness) since 1987 along the New Mexico State Highway and Routes System. Until 2009, NMDOT collected pavement distress data on more than 15,500 lane-miles of pavement in their statewide route system mostly on an annual basis. For at least ten years prior to 2006, the collection of pavement distress data was the responsibility of district construction personnel, proving to be an enormous burden that was taking time away from other tasks. In 2005, NMDOT decided to partner with the University of New Mexico (UNM) and New Mexico State University (NMSU) to handle this massive undertaking. (NMDOT, 2007). Since then, the universities have collected manual data each summer, with a few exceptions. In 2010 and 2011, distress, rutting and roughness data were not collected in New Mexico (Robert S. Young, personal communication, July 2011). The condition of existing pavements is evaluated in two measures: roughness and surface distresses. Combining the two measures, a pavement condition index called Pavement Serviceability Index (PSI) is calculated. This index indicates the overall condition of each pavement section. The NMDOT also uses PSI values to determine the funding eligibility of projects for particular roadway sections.

The Pavement Management section is located within the Project Planning Bureau. It supports the efforts of the department in providing New Mexico with quality highways at minimum cost by providing information necessary to develop cost-effective highway pavement management strategies and to make informed decisions between competing highway projects (NMDOT, 2011). This section of the NMDOT evaluates pavement conditions on a statewide basis and predicts expected pavement deterioration so that pavement preservation, rehabilitation and reconstruction projects can be optimally scheduled.

Prior to 2012, NMDOT procedures for evaluating surface distresses in a pavement section can be found in The NMDOT's Pavement Maintenance Manual, pages 313 and 314. Currently, only a series of criteria are given for each type of distress, which are used by raters performing manual pavement distress evaluations. The raters are to determine the severities of each type of distress and record the extents in a percentage of affected area of the test section. This rating system was developed for use exclusively in the NMDOT Pavement Management System.

1.3 Research Objective

There is only a certain amount of money that can be allocated to projects each fiscal year, making the selection and prioritization important both at the state and federal levels. Each state must assess the conditions of its pavements and pass on the results to FHWA for use on the national scale. Herein lies the problem: Not all of the individual states' pavement management data follow the guidelines that are required by the federal HPMS. New Mexico is one of the states that falls into this category. The codes used by NMDOT in reporting severity and extent do not readily correlate to the data required in the HPMS system. The purpose of evaluating the rating criteria for each type of pavement is to obtain a usable data set for both NMDOT's PMS and the federal HPMS, while maintaining reliability and reducing costs by eliminating the need of having to rate the road system twice.

The purpose of this study is to align the state of New Mexico's pavement management system with that of the federal highway's system, while concurrently reducing variability, increasing data consistency, and providing a greater quantity of specific data items to be considered in pavement preservation while keeping costs under control. To do this, an entirely new data collection protocol has been written and field tested. The three levels of testing include interrater agreement and interrater reliability across the "old" protocol, the "new" protocol, and actual field measurements to judge the accuracy of both old and new methods. By doing this, the accuracy and validity of the pavement distress surveys will be increased. The objectivity and integrity of the NMDOT distress rating criteria for flexible pavements will also be increased through a revision of the old criteria.

A major concern in the implementation of these protocols is the quality of data collected and reported while keeping the time needed to complete these more intensive evaluations at a minimum. The goal of this new protocol is to be able to collect a wider range of data with no additional time spent per milepost than the previous method.

Chapter 2

Background

2.1 Pavement Management

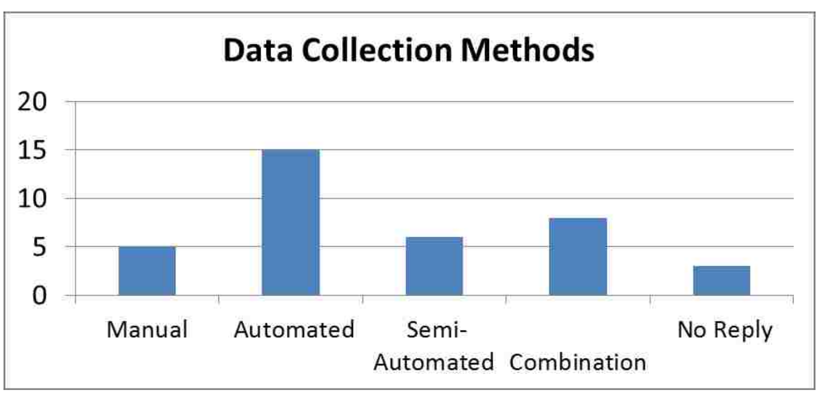
2.1.1 Data Collection

The foundation of any pavement management system is quality data and the use of all available technology in order to obtain and analyze that information. The challenge faced by most government agencies is the reliability and accuracy of data collected throughout the pavement network. State transportation agencies use various methods of pavement data collection. The major methods are manual, semi-automated, and automated collection (Ganesan, 2006). The data collected from these surveys are used to calculate indexes that give a snapshot of pavement conditions across the network, a quantification of distresses found in any particular section. For manual data collection, the pavement raters travel to each site and rate pavement distresses according to their respective agencies' criteria. In a semi-automated method, pavement images are collected digitally using a camera mounted on a van by a process sometimes called *videologging* or simply by 35-millimeter black-and-white photography and are visually rated by a trained person or crew at the state or district office. In a fully automated method, the images are collected and then rated automatically using crack detection software. The level of detail and types of distresses rated depend on the intended use of the data collected. For network level Maintenance and Repair (M&R) indicators, a simple windshield survey may be completed, but for extensive research programs, much more time and effort may have to be involved.

No matter the method, there are inherent flaws in data collection. While each distress is described in detail, and each severity described with specific measurements and photos, there remains a level of subjectivity within the rating process. The accuracy of the field ratings increases as the level of human error is removed: the more automated, the more consistent the data will be. However, a fully automated nationwide system has yet to be implemented, so each state is left to itself to derive the manner of data collection that best fits its needs.

To determine the current state of practice for data collection, surveys were sent to all 50 state transportation agencies and the District of Columbia to compare data collection methods and criteria for individual distresses, specifically related to pavement distress items in flexible pavements. All state DOTs were contacted via telephone and/or email, of which 37 responded to the survey. Table 1 shows the number of respondents using the different types of data collection methods.

Table 1: Number of Respondents using Data Collection Methods



Five agencies reported that they use strictly manual data collection, in which a crew is sent out and follows detailed field guides to obtain the desired item. Most agencies (15) use automated means. Examples of these automated methods include laser profiling vans, where pavement geometry data are analyzed, and digital imaging vehicles, where snapshots are taken at specific intervals while the vehicles are in motion. The primary service companies used are Fugro Roadware Inc. and Pathways Services Inc. The agencies that reported “Automated” data collection rely on the company’s software to produce values to be reported directly to HPMS.

The six agencies that fall into the “Semi-Automated” category use automated means for surveying their pavement systems, but have a designated person or department that manually reviews the images obtained by the vehicles and assigns the required values, particularly for cracking length and percentage of affected area. Eight agencies use a combination of these methods depending on the desired item, and three did not reply to this question primarily because the contact person was not aware of the method used and the person in charge of the state’s data collection could not be reached at the time of the survey.

The five agencies that reported using manual data collection techniques do so in house or hire college students during the summer to perform pavement evaluations. Of the 29 agencies that reported using some type of automated collection methods, 22 are contracted out while seven are agency owned.

The trend is that more and more state agencies are moving toward automated data collection if it is allowed in their budgets. The vehicles and software can be rather expensive, and not entirely accurate as of yet. Hiring college students to work in the summer is an inexpensive way to collect these distress conditions. New Mexico is a state that uses students from the University of New Mexico and New Mexico State University to collect data for both its private PMS and the federal HPMS.

2.1.2 Distress Types

The focus of this research is on flexible asphalt concrete (AC) pavements. There are eight major distresses that the majority of state agencies (including New Mexico) concentrate on in rating flexible pavement. The following descriptions for these distresses are from the 2003 Long Term Pavement Performance program instated by the FHWA (FHWA 2003). The causes of these distresses are taken from the New Mexico Department of Transportation's Pavement Maintenance Manual (NMDOT 2007). Each of these distresses is typically rated according to severity and extent along a test section. These data are then translated into guidelines or recommendations for maintenance, repair, or reconstruction, and are used in pavement serviceability or condition indexes to be used for making other administrative decisions.

Rutting: A rut is a longitudinal surface depression in the wheel path. It may have associated transverse displacement which is called shoving. Rutting is a permanent deformation of any layer due to weakened support layers, poorly compacted layers and unstable wearing surface or overloading. Severe rutting is often caused by excessive asphalt binder in the pavement mixture. Aggregates in these mixtures do not have aggregate-on-aggregate contact so the material flows instead of being locked in place. Rutting is aggravated by hot weather which causes the softening of the asphalt binder.

Raveling/Weathering: Wearing away of the pavement surface caused by the dislodging of aggregate particles and loss of asphalt binder. Raveling ranges from loss of fines to loss of some coarse aggregate and ultimately to a very rough and pitted surface with obvious loss of aggregate. Raveling is caused by oxidation or aging of a paved surface, bad workmanship or materials. Raveling is aggravated by hot and wet weather which causes oxidation and stripping of the asphalt binder.

Bleeding: A buildup of excess bituminous binder found on the pavement surface, usually in the wheel paths. May range from a local discoloration relative to the remainder of the pavement, to a surface that is losing surface texture because of excess asphalt, to a condition where the aggregate may be obscured by excess asphalt with a shiny, glass-like, reflective surface that may be tacky to the touch. Bleeding is usually caused by too much asphalt binder in the pavement mix, excessive prime coat or tack coat or by too low an air void content in the pavement mix. Bleeding is aggravated by hot weather which causes the softening and expansion of the asphalt binder.

Longitudinal Cracking: Cracks predominantly parallel to pavement centerline. Location within the lane (wheel path versus non-wheel path) is significant. If the cracks occur on centerline or outside of wheel path, the cause is usually a poorly constructed paving joint. If in the wheel path, it is caused by excessive deflection due to loading or loss of foundation support probably due to water, insufficient pavement structure or weak support material. Longitudinal cracks within the wheel path are much more serious, and are indicative of early stage fatigue cracking.

Edge Cracking: Applies only to pavements with unpaved shoulders. Crescent-shaped cracks or fairly continuous cracks which intersect the pavement edge and are located within 2 feet of the pavement edge, adjacent to the shoulder. Longitudinal cracks outside of the wheel path and within 2 feet of the pavement edge are included. Edge cracking is caused by loss of foundation support due to water, insufficient pavement structure, weak support material or unstable shoulder.

Alligator/Fatigue Cracking: Occurs in areas subjected to repeated traffic loadings, especially the wheel paths. In early stages of development, it can appear as a series of interconnected cracks. Eventually, it develops into many-sided, sharp-angled pieces, usually less than 1 foot on the longest side, characterized by a chicken wire/alligator skin pattern, in later stages. The primary causes of fatigue cracking are inadequate structural design, poor construction (inadequate compaction), inadequate structural support due to higher than normal traffic loadings, normal loadings on aged and brittle pavement or excessive deflection due to loading or loss of foundation support due to water infiltration, and insufficient pavement structure or weak support material. Small, localized fatigue cracking is indicative of a loss of subgrade support. Large fatigue cracked areas are indicative of general structural failure.

Transverse Cracking: Cracks that are predominantly perpendicular to pavement centerline. These are caused by pavement expansion and contraction due to temperature changes or shrinkage of asphalt binder with age.

Patch Condition: Portion of pavement surface, greater than 4 square inches, that has been removed and replaced or additional material applied to the pavement after original construction. The patches may have been placed for any number of reasons, such as utility work, potholes, or adjacent construction, and evaluated only to determine the intactness of the patch.

2.1.3 Administrative Levels of Pavement Management

Pavement management occurs at two basic administrative levels: project level and network level. Project level pavement management deals with very detailed, technical information centered around one project – a particular segment or section of road. Often, the decisions regarding these types of projects are made by lower or middle management (Peterson, 1987) and involve no consideration of connected sections or the system as a whole. The data required to efficiently manage a single project are those such as foundation strength, materials specifications, climate, and expected number of axle loads. The analysis of these data is much more involved, allowing for the exact diagnosis of structural and material related deficiencies and the corrective action required in eliminating these deficiencies. This method of pavement management is typically dubbed "bottom-up" pavement management, while network (or program) level management is called "top-down" management.

Pavement management at the network level handles policy decisions for the network as a whole holding in highest priority the budget allocation of maintenance and rehabilitation (Haas et al 1994). Decisions in this level are made by upper management and nearly always made for a large number of projects at once for an entire network of highways (Peterson, 1987). This level of management requires skillful planning and evaluation of the system in its entirety in order to make the most practical decisions that will benefit the system, not just one particular area. The network perspective allows the user to properly address the trade-off between heavy rehabilitation and preventive maintenance (Kulkarni & Miller, 2002). Since the level of funding is limited, the option that benefits the network the greatest will be employed, whether it is major rehabilitation for a few segments or preventive maintenance for a large number of segments. Figures 2 and 3 show schematic representations of both top-down and bottom-up management practices, based on the 2002 work of Kulkarni and Miller.

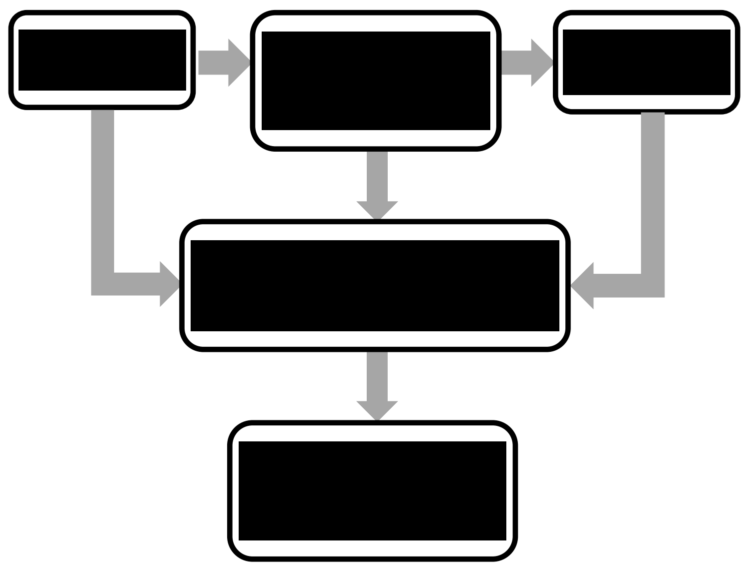


Figure 2: Top-Down Schematic

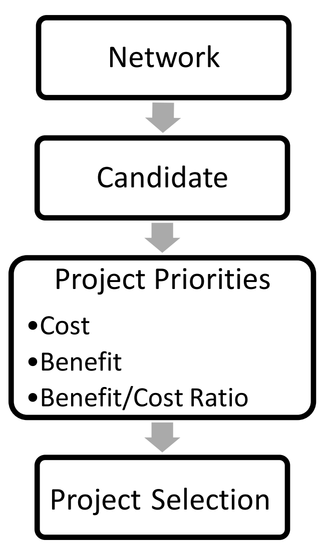


Figure 3: Bottom-Up Schematic

2.1.4 Users of Pavement Management Systems

There are three major groups of users of pavement management systems: technical, administrative, and elected officials. Each group uses the data collected for their respective pavement management systems in different ways. The technical group consists of those engineers who carry out management activities. Generally, these engineers are licensed professional engineers who serve as the point of contact in a management district for the evaluation, preservation, and structural design of pavements. These positions have many responsibilities, including planning activities; participating in design concept conferences; reviewing performance histories of materials; studying processes for pavement construction; maintaining databases for subgrade and pavement material stiffness or structural properties; assessing pavement performance with maintenance staff; and coordinating design strategies for pavement rehabilitation. At the district level, these engineers should be the expert regarding characteristics of the network, such as the evaluation of functional and structural aspects of existing pavements, traffic loading characteristics, prevailing geologic conditions within the district, and the suitability of proposed materials. With the knowledge of these items, the engineer will be able to recommend the most efficient maintenance and rehabilitation processes within his district.

The administrative group of users is generally comprised of planning engineers. The duties of these engineers involves less hands-on tasks and a larger scope of responsibilities, such as performing studies to determine the adequacy of segments of the existing highway system; estimating future traffic demand; determining the sufficiency of the existing system to handle future traffic demands; evaluating the possible improvements; selecting the most feasible improvement and preparing a written report of findings, conclusions and recommendations. Planning engineers often look beyond the scope of roads in their districts and develop comprehensive thoroughfare plans that accommodate present and future load; review and report on projects concerning new shopping center driveways, and median crossover requests; supervise and assist in the collection of land use, population, and vehicle ownership; and travel, economic activity and trip generation data.

The elected officials use the pavement management system to secure funding for new construction, maintenance and repairs of existing infrastructure, and as a means of gaining support from the general public. They must be able to answer questions concerning the effects of lack of available funding and the lowering of standards of a highway network (Haas, 1987) and be able to justify the need for future funding based on historical trends and current condition data.

2.1.5 Maintenance/Repair/Reconstruction

Virtually all roads require maintenance before the end of their service life. Timing of the repair is significant in prolonging the life of the pavement without incurring extravagant agency and user costs. Determining the appropriate time to perform maintenance and repair (M&R) depends on pavement conditions and more recently, life cycle planning. The addition of life cycle cost analysis (LCCA) has helped agencies make more informed decisions as to the timing of M&R schedules. Labi & Sinha (2010) have developed models based on the cost effectiveness of several pavement maintenance schedules, and results show that increasing preventive maintenance is generally associated with increasing cost effectiveness, but only up to a certain optimal point. After this optimal point, the cost effectiveness decreases. It was determined that the position of the optimal point as well as the sensitivity of such cost effectiveness to the preventive maintenance effort are both influenced by functional class and whether user costs are included in the analysis. Condition data must be accurate in order to truly assess the needs of road sections. Historical trends also play into the timing of M&R schedules, as pavements behave differently in different climates and locations.

When considering pavement preservation, the types of treatments fall into three categories, shown in Figure 4.

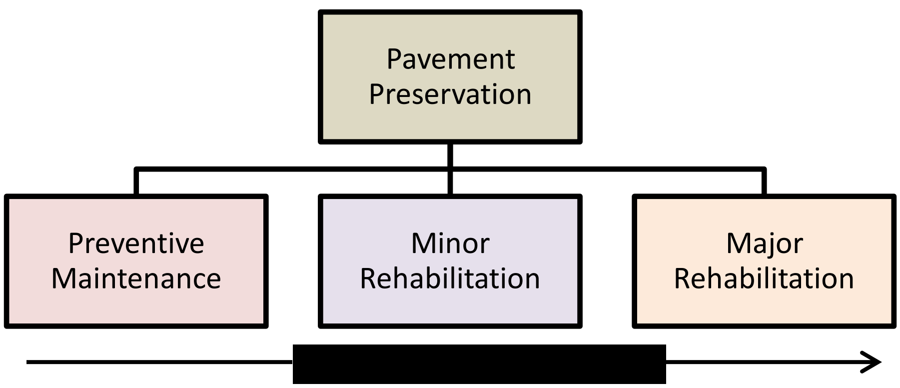


Figure 4: Types of Pavement Preservation

Pavement preservation includes typical routine maintenance activities, such as drainage ditch blading, shoulder grading, and annual operations such as crack filling or patching. It also includes corrective maintenance, which is work performed once a distress becomes severe enough to warrant repair, such as pothole filling or spall repair (Peshkin & Hoerner, 2005).

Preventive maintenance is a key component of an overall pavement preservation program. Sometimes referred to informally as "keeping good roads good," the following definition summarizes the concept of preventive maintenance.

> **Preventive Maintenance**—A planned strategy of cost-effective treatments applied to an existing roadway system and its appurtenances that preserves the system, retards future deterioration, and maintains or improves the functional condition of the system (without increasing the structural capacity) (AASHTO, 1997).

A more specific definition of a preventive maintenance treatment is the following:

> **Preventive Maintenance Treatment**—Any individual maintenance treatment that is used in a preventive manner while not adding any structural capacity to the pavement (AASHTO, 1997).

Examples of preventive maintenance treatments include crack sealing and joint resealing, fog seals, chip seals, slurry seals, microsurfacing, dowel bar retrofitting, diamond grinding, and so on.

The idea behind preventive maintenance is to plan and perform work on roadways before increased levels of effort are required to restore them to acceptable condition. Minor rehabilitation can fall into preventive maintenance depending on the agency. According to the AASHTO Highway Subcommittee on Maintenance, minor rehabilitation consists of non-structural enhancements made to the existing pavement sections to eliminate age-related, top-down surface cracking that develop in flexible pavements due to environmental exposure. Major rehabilitation is not considered part of any pavement management system, as it consists of structural improvements that not only extend service life of the pavement, but load carrying capacity as well.

2.1.5 Pavement Condition Indexes

Traditionally, condition indexes have been used by engineers to describe the current quality of pavement networks and determine maintenance and repair needs and priorities. The monitoring of these indexes over time enables the development of deterioration models, which permit early identification of maintenance and rehabilitation requirements and estimation of future funding needs (Gharaibeh et al 2010). There are several methods of assigning a numeric value to a stretch of pavement in order to indicate its overall condition. The two most commonly used are the Pavement Condition Index (PCI) and the Present Serviceability Index (PSI). There is no consistent method of developing and using these measures, but the most referenced material for these procedures is the AASHO Road Test (AASHO, 1961). Both are based on measurements of roughness, usually International Roughness Index (IRI) data, surface distresses, skid resistance, and deflection. However outdated, this document has served as a basis for several state programs regarding “scoring” a pavement section. This score can quantify the pavement’s performance and be used as a baseline or comparison within a PMS to do one of several things:

Trigger treatment: Once a predetermined score is met, or rather, not met, maintenance or rehabilitation activities can be scheduled.

Determine Extent and Cost of Repair: The score is a numerical representation of the condition of the pavement, so it can be used as an estimate of work that needs to be done and the associated cost.

Determine the Network Condition Index: Scores of pavement sections can be combined to estimate the condition of the entire network.

Allow Comparison of Different Pavements: Since the score accounts for all types of pavement distress, several locations with different problems can be compared on a level playing field.

It is important to note that individual states have their own versions of numerical rating systems, and the index score can vary for a specific pavement condition. Generally, the disagreement among these indexes can be attributed to differences in the distress types considered, weighting factors, and the mathematical forms of the indexes. (Gharaibeh et al 2010). There is an ongoing debate on whether a single index should be used nationwide in order to effectively compare the conditions of roads among several states to obtain a direct comparison rather than assuming the indices are equal (Juang and Amirkhanian, 1992).

2.2 Future Directions of Pavement Management Systems

In the years to come, even greater automation of pavement condition surveys is expected. Equipment and software using the concepts of artificial intelligence and digital imaging are likely to be available to collect data on most pavement distresses including different types of cracks. Global Positioning Systems (GPSs) will be increasingly used in providing location referencing to elements of infrastructure facilities, thus allowing greater and more efficient use of GIS. Another future direction for database applications is Internet or Intranet access to the data and results. Such access will facilitate the use of the data and analysis results by a variety of user groups and agency policy makers and management (Kulkarni & Miller, 2003).

2.3 Statistical Analysis of Data Variability

Data collection is the most scrutinized part of pavement management systems. The quality of data collected is the backbone of such systems, and the data are not always 100% reliable no matter what type of data collection method is used. In the case of manual data collection, the actual rating of distress types and severity levels, no matter how extensive the effort in developing the data collection protocols, still requires subjective interpretations by the evaluator (Rada et al 1997). The subjectivity of the distress ratings may affect the final assessment of the pavement condition and add to the uncertainties inherent to the empirical and statistical pavement performance models and pavement deterioration models, and eventually affect the decision making process of fund allocation and maintenance projects. (Bianchini et al 2010). These subjective inconsistencies can be measured by two indexes: Interrater reliability (IRR) and interrater agreement (IRA). Interrater reliability can be defined as the extent to which two or more parties agree. It addresses the consistency of the implementation of a rating system, and determines whether the rating system requires modification or if more training must be provided to evaluators. Interrater agreement refers to the degree of this consistency.

Several factors can influence the quality of data collected manually, including site conditions, rater bias, even the protocol itself. Site conditions present a unique problem that cannot always be solved. Depending on the time of day, angle of sunlight, rain or other precipitation events, wind, and/or the presence of dirt and dust can affect the outcome of a distress rating, as variability of data is dependent on such issues as season, lighting, surface moisture, surveyor experience and training (Daleiden and Simpson 1998). Smaller cracks that may have been able to be detected due to the angle of the sun may be missed if rated at high noon. Dust may blow in or out of alligator type cracking that may hide or reveal additional features that affect the severity rating.

Some studies have shown consistently that certain raters assign higher or lower severities of a particular distress than others in a group. The inconsistency of differing severities occurs more often than the misinterpretation of distress type and does not appear to consistently have a positive or negative bias (Rada et al 1998). This is indicative of a problem with the rating criteria, or protocol, itself. If raters evaluate the same defect and come to a different conclusion about the severity of the distress, then the criteria may be too vague and need to be rewritten if additional training does not address the problem.

2.3.1 Fully Automated Data Collection

The ideology behind automated data collection is that it can potentially eliminate the two most common problems with manual data collection – rater bias and inconsistency – while also decreasing exposure of agency personnel to accidents. Smith et al (1998) have taken data from Washington and Oregon and compared statistical evidence that automated means of distress rating can achieve the same, if not better, reliability as a standard manual rating. However, the cost increase can be significant to achieve this level of accuracy, and the quality of data and the speed at which it can be collected still need to be improved before widespread adoption of such techniques (NCHRP, 2004).

An automated pavement distress survey system developed at the University of New Mexico in uses an 8-mm camcorder, an inexpensive image-digitizing board, and a 486 personal microcomputer to classify and evaluate pavement distresses. The algorithm is capable of automatically identifying longitudinal, transverse, diagonal, alligator, and map cracking. The program has accuracy in prediction of over 85% in asphalt concrete pavements and over 90% in Portland cement concrete pavements (Chua and Xu, 1994). The described automated survey system is capable of accurately analyzing images captured at a vehicle speed of 15 mph and below. Since then, more advanced equipment has allowed the vehicles to travel much faster and the images captured with higher resolution. The main problem with capturing the images at such high resolution is file size. A tremendous amount of storage is needed for thousands of digital images.

2.3.2 Software Improvements

Fuzzy logic and expert system techniques are effective in evaluating the distresses of flexible pavements while improving accuracy and eliminating subjectivity. A methodology has been developed that uses fuzzy logic for the categorization of distresses that is quicker, faster, and more consistent in classification than traditional manual surveys. An expert system was developed in C language using fuzzy logic for reasoning (Amirkhanian et al 2010). As computing power becomes more advanced and less costly, systems like these may make manual evaluations obsolete.

2.3.3 Geographic Information Systems

A review of practice conducted by the NCHRP in 2004 showed that most DOTs are either currently using or are planning to use GIS or other spatial technologies to support pavement management activities, because enhanced spatial capabilities for data storage, integration, management, and analysis augment many of the PMS functions.

The main improvements that were suggested for facilitating the use of spatial technologies to develop PMS tools included (1) better automatic procedures to facilitate the integration and resolution of data collected and stored using different location-referencing systems, (2) enhanced map-matching techniques, and (3) incorporation of temporal dimension. These enhancements will not only improve a PMS but will also help advance data quality and accessibility throughout the organization, streamlining the work processes.

The principal problems identified with the development and use of GIS-based PMS applications are related to the use of different referencing methods, the level of effort required to develop and maintain the databases, and the handling of temporal issues. Other reported problems included differences among users in the level of detail required to describe the network, accuracy of GPS-collected data when real-time differential correction is not available, excessive user expectations, and the steep learning curve required for users to be able to understand and use the GIS software and procedures. Many of the problems identified relate more to database design and connectivity and PMS application development than to the spatial technologies used.

Chapter 3

Methodology

As stated in Chapter 1, the purpose of this study is to align the state of New Mexico's pavement management system with that of the federal highway's system, while concurrently reducing variability, increasing data consistency, and providing a greater quantity of specific data items to be considered in pavement preservation while keeping costs under control. To do this, an entirely new data collection protocol was developed and field tested. The three levels of testing include inter-rater agreement and inter-rater reliability comparing the "old" protocol, the "new" protocol, and actual field measurements to judge the accuracy of both old and new methods.

3.1 Old Data Collection Protocol

The procedure to be followed in rating a pavement section using the old protocol is very simple. At each evaluation point (e.g., each milepost) walk from the vehicle approximately 530 feet while scanning for distresses, and rate the pavement on the way back to the vehicle. If a distress is present, the raters are to identify the highest severity (Low (1), Medium (2), or High (3)) and the extent of that severity as a percentage of the test section affected: (Low (1-30%), Medium (31-60%) or High (61-100%)) and record it on their field forms. A sample field form is attached as Appendix A, and the evaluation criteria is attached as Appendix B.

In this method, the raters would only note the highest severity present in any particular distress. For example, if several transverse cracks were found all along a test section, and only one crack fell into the High severity criteria, then only that one crack is recorded on the form as a High Severity (3) but Low Extent (1), regardless of the entire section of Low Severity cracks. In this situation, more information is available than is reported, and is one of the downfalls of this existing method. In addition, the old protocol for flexible pavements assumes that the distress of raveling and weathering adopts minimum severity and extent ratings of 1 and 3, respectively, regardless of the road condition, surface type or age of pavement.

The only differences in the old rating procedure that were adopted at the beginning of this study were that Rutting/Shoving was eliminated, due to a request by Robert Young, State Maintenance Engineer of the NMDOT in a letter dated July 27, 2010 stating that rutting would be collected automatically from now on, and Raveling/Weathering could now be rated with a 0 Severity, according to a meeting with the NMDOT technical panel on April 15, 2011.

The old set of rating criteria was also looked at and revised to determine if problems in the language could be contributing to inconsistencies and variations in pavement ratings. The combination of new rating procedures and rating criteria will ultimately reveal the problems with the old system and determine if the revisions made will lead to more consistent results.

Under the old protocol, pavement distress data are collected each year at approximate one-mile intervals on all Interstate, US, NM, Business Loops, and ramps in the New Mexico State Highway System along with other designated routes. NMDOT collects pavement roughness data in-house. NMDOT's Data Reporting Section in the Programs Division collects pavement roughness data on an on-going basis. Pavement roughness data are measured electronically using two K. J. Law Engineers Inc. Model T6600 Inertial Profilometers, each mounted on a Ford E350 van.

The Pavement Serviceability Index is calculated from the Roughness Measure and the Distress Rate. The scale for PSI ranges from 0 through 5, with a value of 5 representing the highest index of serviceability. NMDOT's PSI was devised in accordance with AASHTO's 1986 Guide for Design of Pavement Structures.

The Roughness Measure is the roughness of a section of road measured in terms of the International Roughness Index (IRI) in units of inches per mile. It can range from as low as 10 to nearly 1000.

The Distress Rate is the sum of the severity multiplied by the extent multiplied by the weight factor of each of eight pavement distresses. In mathematical terms:

$$\mathrm{DR} = \sum_{1}^{n}(SRi * ERi * WFi)$$

Where:

DR = Distress Rate of a particular pavement sample.

SR*i* = Severity Rating for the *i*th distress.

ER*i* = Extent Rating for *i*th distress.

WF*i* = Weighting Factor for the *i*th distress.

Here, *i* represents each of the eight distresses that are evaluated in the program; thus, n = 8. Then, the total DR value is the sum of the DR values of each distress (DR*i*). The values of the weighting factors for flexible pavements are given in Table 2 (NMDOT, 2004). These are used to give each distress the effect it has in determining the performance of a pavement.

Table 2: Weighting Factors for Flexible Pavement Distresses (NMDOT, 2004)

Distress Weighting Factor	
Raveling & Weathering	3
Bleeding	2
Rutting & Shoving	14
Longitudinal Cracks	9
Transverse Cracks	12
Alligator Cracks	25
Edge Cracks	3
Patching	2

The distress rate ranges from 0 to 657, 0 being a pavement in pristine condition, and 657 being something representative of a pile of asphalt crumbs. Because the PSI equation relies on more than just the distresses being studied in this thesis, such as pavement smoothness, the overall effect of variability of the estimated distress data will be studied only within the distress rate equation.

3.2 Development of New Data Collection Protocol

The new protocol was developed after researching the needs of the NMDOT and the HPMS reporting requirements. Because certain data items needed to be reported in specific ways, i.e. total feet per mile of transverse cracking and percent area affected by alligator cracking, the data collection process needed to be altered in a way that reflected these new requirements. The old method of having the students collect data only by severity and extent does not effectively translate into quantitative measures and no correlation was found while examining the data sets. Further explanation of the development of the new protocol is discussed under each HPMS reporting requirement later in this section.

The new protocol suggested in this report follows the needs of HPMS data reporting as well as the revised needs of the NMDOTs pavement management system. In the aforementioned letter sent by Robert Young, a few more changes were made to data required by NMDOT. Rutting and shoving are to be collected using an automated system in the future and patching is to be eliminated entirely. The main reasons for this change are safety of the pavement evaluating crew and the ease of data collection by an automated system already in use for IRI. In addition, longitudinal cracking occurring in the wheelpath is to be combined with alligator cracking, as these phenomena are both caused by cyclic loading of pavements, and longitudinal cracking outside the wheelpath is to be combined with edge cracking. The most significant change is that each severity within each distress needs to be rated to provide a more accurate picture of the conditions of pavements in New Mexico. Previously, “severity ruled extent” – that is, only the worst severity was to be reported.

The 2010 HPMS Field Manual provides detailed descriptions of each distress, photos of example sections, and sample methods of data collection in order to achieve consistent results over all state agencies. This is the primary reference used to develop these new field rating procedures. In summer 2010, a survey was conducted regarding data collection methods that other states use. Of the states that participated in the survey, several submitted their rating manuals that they use for manual data collection. These were also consulted when these new procedures were created.

By reviewing NMDOT's old criteria and simplifying some of the descriptions for pavement distresses, new criteria were written that should make identification of distresses in the field a bit easier on the pavement evaluators, and save time and therefore save money. The example of the new criteria that evaluators carry with them in the field is attached as Appendix C.

Every method detailed here assumes manual collection of data from the roadside. In training, the pavement raters will calibrate their paces to be able to accurately measure lengths of distresses without actually having to measure them. This will save time and improve safety by allowing the rater to stay out of traffic lanes, which also avoids costly and inconvenient road closures.

The procedures in place for collecting data in the field were modified to include HPMS reporting requirements as well as handle the needs of the NMDOT's PMS. For flexible pavements, HPMS only requires three items to be reported:

- Rutting (item # 50)
- Fatigue Cracking (or Alligator Cracking – item # 52)
- Transverse Cracking (item # 53).

NMDOT requires data collection for the following items for flexible pavements:

- Raveling/Weathering
- Bleeding

- Rutting and Shoving
- Longitudinal Cracking
- Transverse Cracking
- Alligator Cracking

As previously mentioned, rutting will be evaluated automatically, effectively removing it from this list. The concentration will be on fatigue and transverse cracking for the HPMS system, and the other items required by the NMDOT will be reviewed in order to simplify the process. The following sections explain the new protocol in further detail.

HPMS Item # 52 – Cracking Percent (See page 4-92 of the HPMS Field Manual)

This item represents fatigue cracking and longitudinal cracking located within the wheelpath. HPMS requires a percentage of total sectional area affected by this type of distress be reported to the nearest 5%. In order to ensure any instance of this distress is captured, the percent area will be rounded up. This will take care of the instance that a small area of the sample section is affected that would otherwise round down to a zero extent, since alligator cracking is a major indicator of pavement distress. To obtain the data, the pavement rater will "pace off" the lengths within the section that display this sort of distress, recording the length and approximate width of the distress on their field forms. The approximate width is recorded by counting the number of wheelpaths that the distress occurs in, either a 1 or a 2. The HPMS Field Manual assumes that this type of cracking only appears in the wheelpath and that we can assume a 2 foot width for each wheelpath. Each severity will be rated for use in the NMDOT PMS.

HPMS Item # 53 – Cracking Length (See page 4-98 of the HPMS Field Manual)

This item refers to transverse cracking. HPMS wants this item reported in linear feet per mile. The rater will simply count the number of cracks that are at least six feet long (half a crack counts as a whole crack), within each severity level and record the totals on their rating forms. NMDOT wants each severity reported for use in its PMS, however, the *total* number of cracks across all severities will be used for HPMS reporting, as severity is not considered.

Additional NMDOT Requirements

Because Raveling/Weathering and Bleeding have two of the lowest weights in the state's old PSI formula and because they are the most difficult distresses to evaluate consistently, it is recommended that these two items be collected on a Present/Not Present basis. Since it is assumed that raveling and weathering will occur over an entire section, the raters would only have to note the severity of the distress, and that bleeding occurs in spot locations, the raters would have to note each severity they find. In order to validate these assumptions and justify changing the data collection procedure, historic data were analyzed. New Mexico State University (NMSU) performed an analysis of the data recorded from 2006 to 2009 from the pavement evaluation crews that were sent out every summer. This data set includes every instance of every distress at every milepost recorded by The University of New Mexico (UNM) and NMSU. The basic results of the data analysis are as follows:

- The extent of 3 (61-100%) for Raveling/Weathering occurred more than 87% of the time, indicating that this distress occurs over the entire rated section most often. This makes sense, as this distress is usually caused by a mix issue or poor construction quality and occurs throughout the batch. See Figure 5 for further clarification.

- The extent of 0 (None found) for Bleeding is the most common rated, over 65% of the time. If bleeding is present at all, the extent is only a 1 (0-30%) 22% of the time, indicating that bleeding affects large areas of the test section very rarely. Since bleeding is a buildup of excess asphalt binder, it is most commonly found in patches where the binder has seeped up onto the surface in hot weather. See Figure 6 for further clarification.

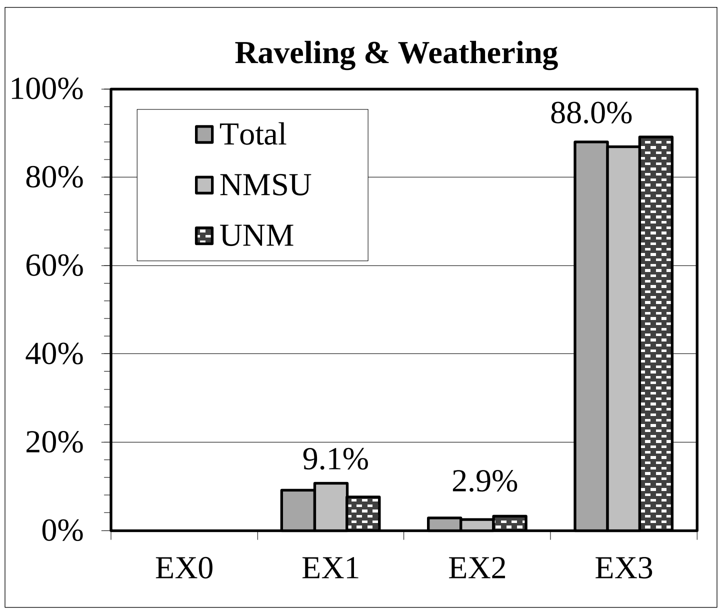


Figure 5: Average Extent of Raveling & Weathering

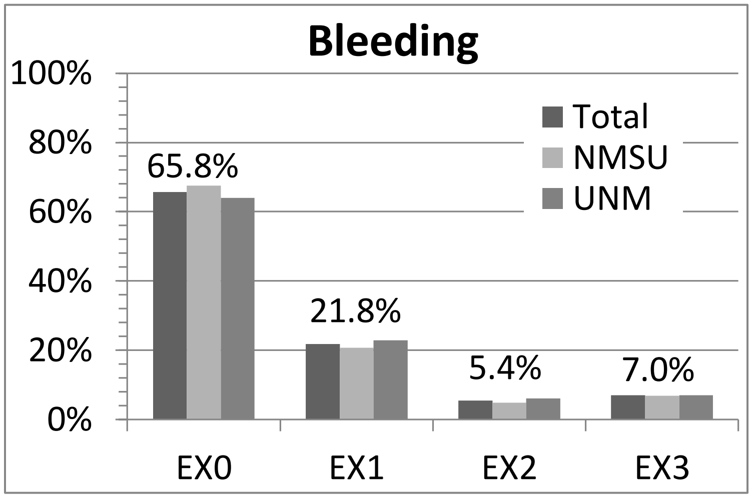


Figure 6: Average Extent of Bleeding

Additionally, Oregon DOT uses a similar rating system for Raveling/Weathering, a simple Yes/No, noting the severity. These results of the data analysis and the presence of another state practicing these methods validate that we may be able to utilize a Present/Not Present rating system in the future for Raveling & Weathering. Additional reasons for wanting this change is that these two distresses are indeed the hardest for pavement raters to accurately identify and the weights in NMDOT's Distress Rate and PSI equation are so low that any major change made here affects the overall values very little. At this point, it has been determined that the minimal effect of these distresses far outweighs the need to evaluate these distresses in great detail. Accurate information can still be provided to the NMDOT without spending the additional time and money required to rate these distresses extensively.

Raveling and Weathering. This item is not needed for HPMS. NMDOT still needs it reported for its PMS. Based on the data analysis performed by Dr. Bandini of NMSU and the Oregon pavement evaluation manual, we are confident that if these distresses exist, they will most likely be present along the entire test section. The only required action from the pavement raters would be to indicate the severity of the distress on their field forms.

Bleeding. This item is not required for HPMS. NMDOT still needs it reported for its PMS. We are suggesting collecting data on a Present/Not Present schedule due to the low effect on the PSI of the test section and the fact that historically, the largest extent reported is 1 (0-30%). When collecting data in this manner, the pavement evaluator will simply note the severity of any bleeding that occurs within the test section, assuming an extent of 1.

Longitudinal Cracking. This item is not required for HPMS. NMDOT still needs it reported for its PMS. This distress refers to longitudinal cracking outside the wheelpath, located anywhere within the test section. This distress will be evaluated the way it has always been done, except the rater will evaluate each crack in terms of severity and extent.

The top three distresses on the field form are Raveling/Weathering, Bleeding, and Alligator Cracking. The pavement evaluators should focus on these three distresses on the way "out," that is, away from the vehicle. The raters should be able to easily and quickly evaluate raveling and weathering and indicate the worst severity on their field forms. This leaves the trip out to concentrate on alligator cracking, which has a weight of 25. Instances of bleeding can be noted effortlessly because only severities need to be marked.

The last two sections on the field form are to be collected on the return trip to the vehicle: Transverse Cracking and Longitudinal Cracking. Because Edge Cracking and Longitudinal Cracking are now combined into a single category and not needed for HPMS data reporting, the concentration will be on transverse cracking, which carries a weight of 12. As the raters walk back to the vehicle, they can simply count the number of cracks that occur within each severity. The old method for obtaining a weight for the new category of longitudinal cracking is adequate for these requirements.

By separating the distresses into two different time frames, it is believed to take some guesswork out of the evaluation process. Raters will not be overwhelmed by roads in bad condition that display several distresses if they are to concentrate on only a few at a time.

3.3 Testing of Old versus New Data Collection Protocols

Two separate protocols were evaluated in this study – NMDOT's old protocol and the new protocol developed for this study. The purpose of the comparison was to determine if the new protocol performed as well or better than the old protocol in terms of variability and accuracy.

3.3.1 Initial Data Collection

Field data were initially collected using each of the two protocols. These data were then used in each of the three tests listed above. To collect the data, two separate trips (one for the old protocol and one for the new) were made at least one week apart. The time between ratings was to ensure the rater could not remember what he recorded so that each protocol could be fairly evaluated, making sure the data are not biased and that they can be compared for similarities and differences across time.

Two graduate students worked on this project, without having gone through any previous training. The additional students who were chosen to participate in this project had previously worked for the NMDOT summer pavement evaluation project. This was done so that training could be kept to a minimum, as the experienced raters already knew the distresses and only a refresher course was done to reinforce safety expectations and review the process.

In order to achieve the level of data quality needed to validate the new protocol, it was decided to base the data on the following report: A Framework for Assessing and Improving Quality of Data from Visual Evaluation of Asset Conditions, (Cordova 2010). In this project, data points were monitored for quality by two measures: interrater agreement – the ability for a group of pavement raters to arrive at the same result - and agreement over time – the ability of the same rater to provide the same evaluation across a given time. Twenty four of the same mileposts were used in evaluating this new protocol that were used in the previous study as a means of comparison. This combination of mileposts was chosen based on the previous data which showed that each severity could be found in each distress. The mileposts used for this project are listed in Table 3.

Table 3: List of Mileposts

Route	Milepost				Direction
NM0041	0	1	2	3	P
NM0041	29	30	31	32	P
NM0006	0*	1	2	3	P
NM0556	12	13	14	15	M
NM0014	0	1	2	3	P
US0550	0	1	2	3	P

*The legal definition of Milepost 0 was not used. Milepost 0, according to the NMDOT's Legal Definitions file is north of the I-40 off-ramp. This piece of roadway is not traveled; it ends in dirt, and is literally crumbling with weeds and grasses growing through the cracks. A managerial decision was made to use the first 1/10 mile south of the off-ramp to get more realistic distresses to be used in comparing evaluation methods.

The approaches to the pavement section and safety procedures are the same in both the old and new protocols. The main points are discussed below, as taken from the 2009 UNM Pavement Evaluation Report (UNM 2009):

1) A crew composed of two people must perform the visual surveys. One will serve as the distress rater while the other will serve as the safety person (also referred as safety spotter) to watch for hazards on and off the road. The two crew members will take turns in both roles. While in the section, the safety spotter should alert the rater of any hazard and should alert the traveling public of their presence.

2) Approach mile marker where survey location is to begin; anticipate this location because you will have to slow to a stop at the milepost (MP). Approximately one-half mile from the MP, turn on your emergency Light Bar and right hand turn signal. Slow gradually to a stop well off of the pavement adjacent to the MP (for positive direction). For minus direction, park 0.1 mile before MP. Turn on the emergency flashers on the vehicle (Hazard Lights; 4-way). Leave vehicle running (power for light bar).

3) Safely exit the vehicle looking for traffic from the rear. Put on required NMDOT safety vest and cap. Install Survey Crew sign on rear of Pavement Evaluation Vehicle (secure firmly using straps). Obtain all necessary safety and evaluation equipment for conducting survey.

4) The Pavement Evaluator (PE) evaluating the pavement should have the following equipment: Clip Board & Pen, Evaluation Form, Rut Bar (Level), Steel Rule. The Safety Spotter (SS) should have the following equipment: Slow/Slow Sign w/ Strobe, and any necessary equipment to mark off 530 feet and to locate evaluation points.

5) Mark off ~530 feet to the front of the vehicle (with traffic flow). As you are marking off the distance, the PE can be doing a quick initial evaluation from the side of the road.

6) Upon marking off the required 530 feet, the two-person crew shall return toward the parked vehicle. The PE shall evaluate the pavement as required, while the SS watches for traffic and advises the PE of adverse conditions that may imperil either of their safeties. The SS shall use the SLOW/SLOW sign and warning strobe as necessary to warn traffic. The SS shall continuously monitor oncoming traffic as the team returns toward the parked vehicle and remain along the edge of the road with the SLOW/SLOW sign facing

traffic and between the SS and the travel lane. The SS shall stay even with the PE. In rare cases where there is limited sight distance, the SS may position themselves further up the road (toward the oncoming traffic) to improve their view of oncoming traffic. In no instance shall the SS be outside of voice range of the PE.

7) Upon return to parked vehicle, store all measurement and safety gear. Store data form in accordance with standard Data Quality Management procedures. Secure safety belts and slowly move down the shoulder to the next milepost. If the shoulder is not wide enough, then move in to the traffic lane safely and drive to the next milepost to repeat the testing protocol. Remember to use flashing hazard lights until up to speed.

8) At the end of a major test section and/or at the end of the day, the Survey Crew sign must be removed and stored and Light Bar & Hazard Lights must be turned off. 8) At end of workday, all equipment must be properly accounted for and stowed in motel and or vehicle trunk. Account for all equipment at beginning of each workday. Check vehicle per daily vehicle check list each morning.

3.3.2 Interrater Agreement

There are currently several methods available for testing interrater agreement; however, one of the simplest and most robust is the average deviation (AD) index. As explained by Burke & Dunlap (2002), the AD index is actually a measure of disagreement, such that a value of zero means that there is zero disagreement, or, total agreement. This measure was developed for use with multiple evaluators rating a single target on a variable using an interval scale of measurement. This index estimates agreement in the metric of the original scale of the item (i.e., it has the same units as the item targeted) and therefore can be considered a pragmatic measure (Burke et al 1999). The AD index may be estimated around the mean (AD_M) or median (AD_{Md}) for a group of evaluators rating a single target, such as a stretch of pavement on a single item such as pavement distress. Each rating is compared to the others in a group, and the deviation from the median is calculated, giving a relative "distance" from the expected value.

AD_{Md} values are computed as follows:

$$AD_{Md(j)} = \frac{\sum_{n=1}^{N} |xjk - Mdj|}{N}$$

Where *ADMd(j)* is the average deviation from the median computed for an item *j*, *N* is the number of judges or observations (consequently the total number of deviations for an item), *xjk* is the kth judge's score on item *j*, and *Mdj* is the median for item *j*.

The scale *ADMd(J)* is then computed as:

$$AD_{Md(J)} = \frac{\sum_{j=1}^{J} ADMd(j)}{J}$$

Where *ADMd(J)* is the average deviation computed from the median for *J* essentially parallel items and *ADMd(j)* is defined as above. Although *ADM* and *ADMd* scale values can be computed directly from respective scale means and medians, these latter values are based on composite scores and cannot be directly interpreted in terms of response options or units of the original measurement scale (Burke et al 1999).

Since there is rarely total agreement among evaluators, a cut-off value of ***c/6*** can be used to determine whether there is a consensus among evaluators, where ***c*** represents the number of response options. This ***c/6*** was developed by assuming 0.7 as a lower cut-off limit and rearranging the correlation coefficient, selecting a uniform distribution for the likelihood of an inexperienced rater choosing any possible value from the set and adjusting the results for average deviation (Cordova, 2010). Values lower than the cut-off point mean acceptable levels of consensus, while a value that falls over the cut-off point would indicate a problem of consensus between evaluators.

Following the old protocol, only the highest severity of each distress found in a test section is reported along with its extent. For every distress, the range of values that is available for selection in the old protocol is 0 through 3, giving 4 choices. Therefore, ***c*** is equal to 4, and the cut-off value is 0.67. The smaller the deviation from the median, the better, and any AD index above 0.67 is considered a problem, and the underlying issues must be resolved to correct it.

The new protocol involves changing the way the data are collected for almost every distress encountered. Therefore, a new analysis was required for each distress. The method used is the same, the average deviation about the median, but certain values were adapted based on the format that the data were to be reported in. In the new protocol, there are now more choices and more severities reported, leading to the obvious need for rewriting the analysis program. It is important to note that the reported values for each distress need not be exact among raters – the main point is to assess whether a distress is present and attempt the most accurate evaluation possible. If all ratings are similar, then the evaluation has succeeded in giving a valid reference point for the general condition of the test sections. Each distress will be explained in detail below.

Raveling/Weathering. This data item requires that only the worst severity be reported, as it is assumed that this distress affects the entire section based on the cause of the problem (giving it an automatic Extent 3). The analysis for this item is simple: compare the values reported from each rater, since the only available options are 0, 1, 2, and 3. The number of alternatives will be 4, and the cut-off coefficient will remain 0.67 for this distress.

Bleeding. Bleeding will be evaluated in a fashion similar to raveling/weathering. Since more than one severity can be reported within a test section, the sum of the observed severities will be compared among raters. For example, if one rater finds severities 1 and 2 on a test section, the sum will be 3. If another rater finds severity 2 only, the sum is 2. The possible numbers reported are 0, 1, 2, and 3, in a combination of none or all severities, leaving the number of alternatives equal to 7: 0, 1, 2, 3, 4, 5, and 6. The cut-off coefficient for this distress will be 1.17.

Alligator Cracking. Alligator cracking will be evaluated by pacing off the lengths of each severity located within the test section. In order for this value to be consistent among different raters, it will be converted to a percent area of the section (each rater will most likely have a different pace length). The percent area will be rounded to the nearest 5%, according to HPMS requirements, and compared among each rater. The number of alternatives will be 36, because the highest possible area within a test section will be 33.3%, and, due to the rounding, will be reported as 35%. It was decided to use the range of 0-35% as opposed to multiples of 5 (i.e. 0, 5, 10, 15, 20…etc.) to avoid an exaggeration of 5 times the actual deviation which would skew the results negatively. The cut-off value here will be 6.0.

Transverse Cracking. Transverse cracking is evaluated by counting the number of cracks that occur within a test section. HPMS requires this data item be reported in linear feet per mile, which effectively multiplies the number of cracks times 12 feet (the average assumed lane width) to obtain a length, and again by 10 to convert the 1/10 mile sections that are evaluated to a full one mile section. Hence, the number reported to HPMS will be a multiple of 120. According to the state of Nevada's Pavement Distress Manual, the most severe case of transverse cracking will have a crack occurring every 5 feet, which works out to 106 cracks in a tenth of a mile test section, if the maximum number of cracks occur. See Table 4: Nevada Transverse Cracking Limits.

Table 4: Nevada Transverse Cracking Limits.

NMDOT Extent	Minimum # of Cracks	Maximum # of Cracks
0	0	0
1	1	35
2	36	71
3	72	106

Therefore, the number of alternatives will be 12,721 (106*120 = 12720, plus one for the zero option.) Again, it was decided to use a continuous range rather than multiples of 20 to avoid skewing results. The cut-off value in this case is 2120.2. After further evaluation, these numbers were determined to be grossly overestimating the values the pavement evaluators would ever obtain and the cut-off value for good agreement ended up being so large that any variability was dramatically reduced. Data from 2006 to 2009 was analyzed in order to determine the largest number of cracks that have been recorded in a test section in order to more realistically set these limits. This limit was found to be 51. This changes the number of alternatives to 6120 and the cut-off value for good agreement to 1020.0.

Longitudinal Cracking. Longitudinal cracking is reported in each severity along with corresponding extents. In order to obtain a single index value to be used for comparison, it has been decided that a level of distress should be calculated. This level of distress consists of the sum of each severity multiplied by the extent of that severity. This distress is not required by HPMS, so it will be used entirely for NMDOT purposes. The number of alternatives is the sum of each possibility of combinations, 0 – 18, giving 19 total possibilities. The cut-off value will be 3.17.

3.3.3 Detailed Field Measurements

The new protocol was designed to collect reasonably accurate data using pavement evaluators on the roadside. This decision was based predominantly on safety concerns, but also factors like time needed for collecting the data and the availability of advanced technology, i.e. the automated rutting collection process, which can be done much faster. The preceding analyses have demonstrated that a group of pavement evaluators can be trained relatively quickly to produce ratings that are both consistent among the group and over time; however, for the results to have meaning, they need to represent the actual conditions of the pavement. These detailed measurements of pavement distress were conducted to ensure that the data being estimated by the pavement evaluators are accurate for HPMS data reporting. If the estimating process is not accurately portraying the distress on the roadways, a correction factor can be applied to more closely match the pavement conditions found (Bandini et al, 2012).

Remember that, according to the new protocol for visual distress surveys, all severities and their corresponding extents are rated for each distress type. The distress data obtained during the visual surveys for flexible pavements include the following:

1. <u>Alligator cracking:</u> For each severity, the area of alligator cracking is assessed from the rater's pace count, for one or two wheel paths, along the parts of the section that exhibit this distress. Note that pace counts for longitudinal cracks along the wheel path(s) are included in the area of alligator cracking of severity 1 (Low). The rater reports the pace count, not an extent rating, for each severity level. The rater's pace count can be converted into extent rating using simple arithmetic and recommendations from the HPMS Field Manual.

2. <u>Transverse cracking</u>: For each severity, the rater's count of transverse cracks (equal to or longer than 6 ft) is recorded and not the extent rating. The rater's crack count can be converted into extent rating, again, using simple arithmetic and recommendations from the HPMS Field Manual.

In order to validate the procedure, the approach was to find correlations through statistical analysis between raters' data and detailed field measurements of the extents of these two distress types on the same pavement

s. The detailed measurements in test sites were performed under the supervision of Dr. Paola Bandini of New Mexico State University as part of project NM10MNT-01 (Bandini et al 2012). For these tests, traffic control and warning signs were provided by

NMDOT District 1 Maintenance Crew based in Las Cruces, NM to ensure safety of the raters, research assistants and traveling public during the detailed field measurements. No traffic was allowed in the test section during the work.

As detailed in the project report, fifteen test sites of flexible pavement were selected to include interstate highways (high traffic volume, heavy traffic), U.S. highways (medium/high traffic volume) and New Mexico highways (thin AC layer, low traffic volume). In selecting the test sites, important factors were considered, including their proximity to NMSU main campus, pavement surface condition (distress types, severity and extent), absence of potential road hazards for the pavement evaluators, and possibility of minimizing disruption of traffic and access to adjacent roads and private property during the fieldwork. Table 9 lists the 15 test sites selected for this work and their location. For the purpose of these measurements, a test section was 0.1 mile long and had the width of the driving lane, 12 feet. These are the same characteristics of the sample sites evaluated in the NMDOT's Pavement Distress Data Collection Program

Table 5: Test Sites for Field Measurements of Fatigue and Transverse Cracking, (NM10MNT-01, Bandini et al 2012)

Route	Milepost		Direction	General Location
	Begin	End		
I-25	15.0	15.1	Positive (Northbound)	About 7 miles north of city limits, Las Cruces, NM
	15.1	15.2		
	15.2	15.3		
I-10	129.0	129.1	Positive (Eastbound)	About 3 miles west of Las Cruces Airport Interchange
	129.1	129.2		
	129.2	129.3		
	129.3	129.4		
US 70	144.9	144.8	Minus (Westbound)	Picacho Hills area, Las Cruces, NM
	145.0	144.9		
NM 28	17.0	17.1	Positive (Northbound)	About 11 miles south of junction with NM 101 (University Ave.), Las Cruces, NM
	17.1	17.2		
	17.2	17.3		
NM 478	19.0	19.1	Positive (Northbound)	About 2 miles south of junction with NM 373 (Union Ave.), Las Cruces, NM
	19.1	19.2		
	19.2	19.3		

Three experienced pavement evaluators of the NMSU team performed independent ratings using the new protocol of the selected test sites prior to the detailed measurements,. After these ratings were completed, the evaluators, graduate research assistants, and their professor collaborated and determined by consensus the severity and extent of each distress.

Actual crack lengths of transverse cracks were measured using a measuring wheel. The actual area affected by alligator cracking was measured using 3 foot by 3 foot grids made of thin laths.

Measurement Procedures

The pavement evaluators and research assistants followed these steps to carry out the measurement of areas of alligator cracking and length of transverse and longitudinal cracks in each test section (NM10MNT-01, Bandini et al 2012):

1. Using white chalk and a measuring wheel, mark the beginning and end of the test section. Also, mark 50 ft long subsections starting from the beginning of the test section. (The last subsection will be shorter, being just 28 ft.) Next to each chalk mark, write the cumulative distance from the start of the section to that point; for example, 0 feet, 50 feet, 100 feet, 150 feet, and so on. Make these marks on the pavement surface near the edge stripe.

In each 50-ft long subsection:

2. Identify the areas of alligator cracking, transverse cracks and longitudinal cracks. Using chalk, write “A,” “T” or “L” on or near the distress. (“A” stands for alligator cracking, “T” stands for transverse cracking, and “L” stands for longitudinal cracking.) Write the corresponding severity rating (1, 2, or 3) next to the letter. For example, “A2” means alligator cracking of severity 2. Use a ruler to measure the crack width if needed to determine severity level. For transverse and longitudinal cracks, draw a chalk mark across the defect indicating the start and end of each crack to facilitate the length measurements (Figure 7).

Figure 7: Chalk Markings in Preparation for Measurements (Dr. Paola Bandini, NMSU)

3. Using the measuring wheel, determine the ***length*** of each transverse crack, regardless of their length or severity. Record the severity and lengths of each crack. Take into account the following:

    - Do not include or measure the part(s) of a crack that lie(s) outside the test section or beyond one foot of the mid-lane stripe or edge stripe.

    - Measure two types of length for each crack. First, measure the "*actual crack length*" following the crack's wavy or sinuous profile (Represented by the continuous line in Figure 31). Record this length, reset the counter of the measuring wheel, and measure the "*length*" as if the crack were composed entirely of a straight segment. For the 2-point length, follow an imaginary straight line between *two points* located at the ends of the crack (Represented by the segmented line in Figure 8). Record this length and reset the counter. Record the corresponding crack type and severity.

- After each measurement and using chalk, draw a line across the crack indicating that it has been already measured and recorded.

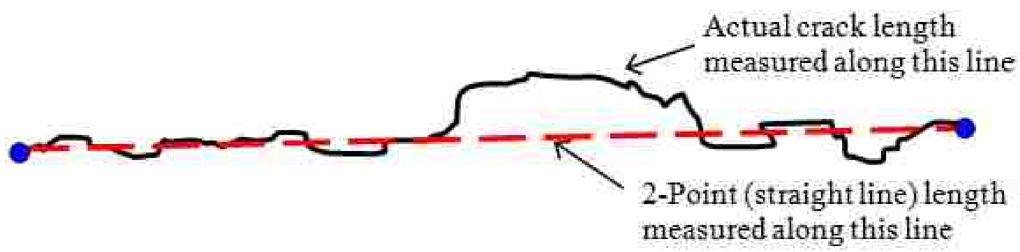


Figure 8: Actual vs. Straight-line Length of Transverse Cracks (Dr. Paola Bandini, NMSU)

4. Using chalk, draw an outline around each alligator cracking area. The outline should fully contain the alligator cracking area within the test section (Figure 9). This line should approximately follow the perimeter of the alligator cracking area, about 6 inches beyond the outer cracks. (Alligator cracking must contain at least three connected cells; otherwise, disregard it.)

5. Place the reference grid on the pavement surface over the outlined alligator cracking. Mark the location of the grid's corners with chalk as shown in Figure 10. Count the number of whole and partial "squares" of the grid that lie on the outlined alligator cracking area as shown in Figures 11 and 12. Record this number with the corresponding severity.

6. If the grid is smaller than the outlined area, move the grid and count the remaining "squares" until the complete distress area has been covered. Using chalk, mark the alligator crack indicating that it has been already measured. Compute the surface area as the sum of the number of whole and partial squares times the pre-determined area of each square.

7. If the alligator area is approximately rectangular, you can use a measuring tape to determine the dimensions and calculate the area.

8. Compute the following for each 50-ft subsection and for each severity:

    - Cumulative area of alligator cracking,
    - Cumulative "actual length" of transverse cracks,
    - Cumulative "2-point length" of transverse cracks,

In the field test, three experienced raters were used. Their averaged data were considered instead of their individual data.

Figure 9: Outlining Cracking Area

Figure 10: Grid Application

Figures 9 and 10, Dr. Bandini, NMSU

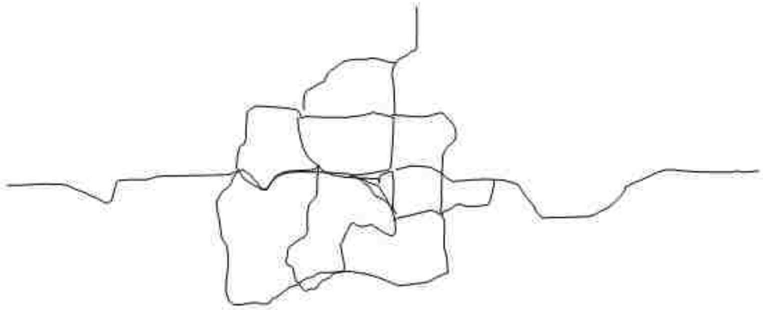


Figure 11: Alligator Cracking Instance

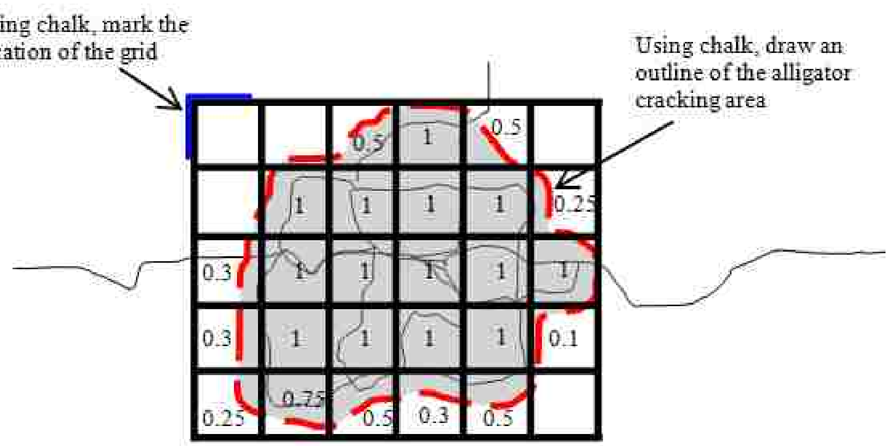


Figure 12: Measuring Alligator Cracking Using Grid

Figures 11 and 12, Dr. Bandini, NMSU

Chapter 4

Results and Discussion

4.1 Interrater Agreement Results

4.1.1 Interrater Agreement - Old Method

Four evaluators were used in collecting data in the old method. Each rater independently evaluated each milepost to obtain data points to be used for comparison. Ideally, the differences in average deviations about the median would be 0, but anything under 0.67 is considered acceptable.

Figures, 13 and 14, show the average deviation about the median for each distress in both rounds of evaluations, separated by distress severity and extent. Note: the red line represents 0.67 – the cut-off value for agreement. The horizontal axis shows the distress, while the vertical axis shows the average deviation about the median.

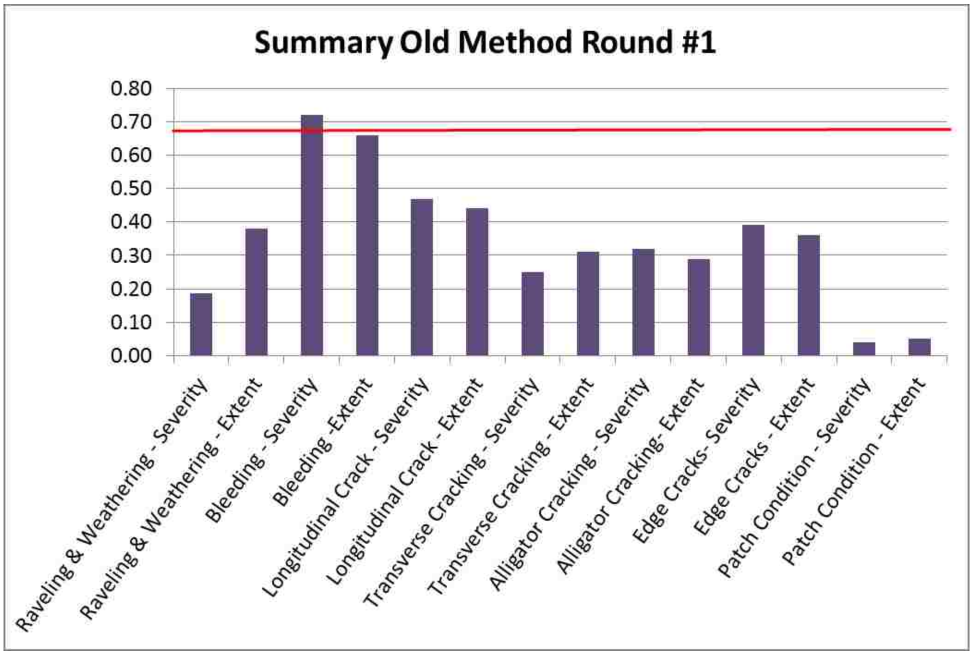


Figure 13: Summary of Round 1

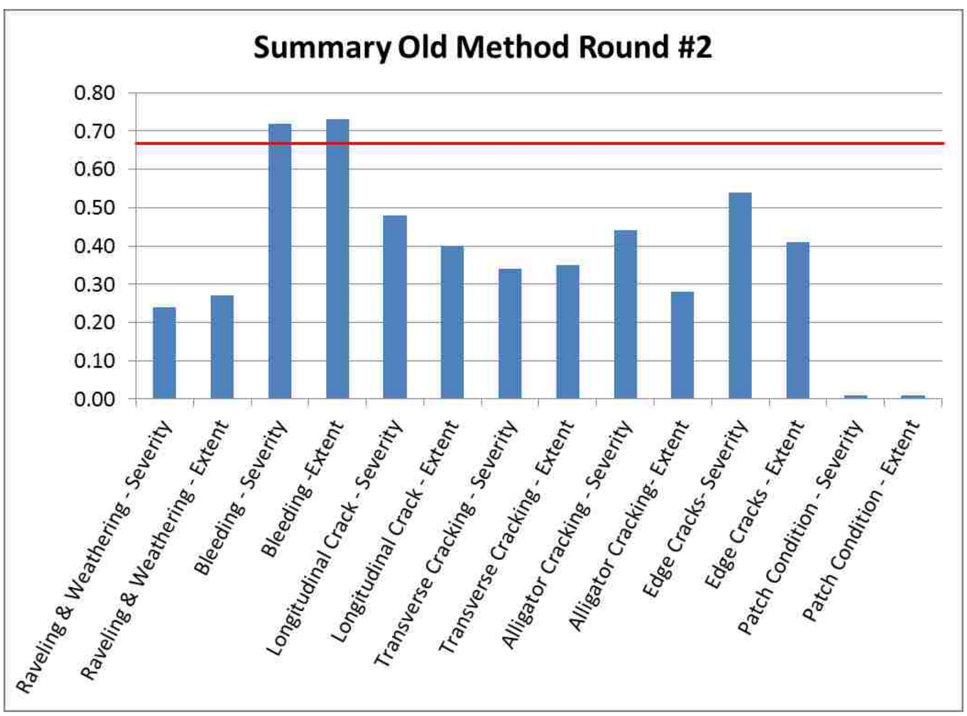


Figure 14: Summary of Round 2

It is worthwhile to note that all distresses may be shown on the same graph because they are all rated in the same manner, with both severity and extent having a range of 0-3 and same number of alternative choices. The general shape of the two graphs is consistent, even though the values are slightly different, which demonstrates stability in the rating methods. Overall, the agreement between raters is good, with only one distress having values above the cut-off – bleeding severity and extent. Previous research has been done showing that bleeding had the greatest variability overall, both among student raters and between student raters and expert raters (Bogus et al. 2010). An initial review of data from the NMDOT pavement evaluation program in 2006 shows that bleeding is a distress that is difficult for inexperienced raters to evaluate, while alligator cracking is a distress that is both repeatable and reproducible. (Bogus et al 2010), and both of these trends have been verified through this experiment as well. These findings further illustrate the need for revision of the evaluation protocol and criteria for bleeding.

4.1.2 Interrater Agreement – New Method

The following sets of analyses were performed across raters to determine whether the criteria and protocols written could be successfully applied to a group of raters and obtain a set of results that are relatively identical. This set of tests is examined by milepost – the ratings that each evaluator reported are compared according to rating criteria.

Each distress evaluated under the new method must be shown on a separate graph due to inconsistencies in ranges of scores and numbers of alternatives. However, the same basic methods of analysis are used and the general approach remains the same. Additionally, since the analyses performed on this data set have not been done before, and the possibility of higher variation, it was decided that two indices be compared to prove agreement – the average deviations among the mean and the median. After evaluation using both of these measures, the values are close enough to each other that there is no real effect on the statistical significance of either measure, but both are shown for clarity.

Alligator Cracking

Several analyses were carried out because of the new requirements set forth by the NMDOT. All severities are mandatory in the new protocol, so each severity was evaluated by itself to determine whether the criteria describing each severity were adequate and could be repeated. An analysis was also performed on the total alligator cracking reported per milepost as another test of the evaluation procedure. If the individual severities of alligator cracking rated in the field had large variability and the total did not, it would indicate a problem with severity criteria. If the total alligator cracking had large variability, it would indicate a problem with rating the extent of the distress. The following figures show the results from the data analysis. In all cases, the cutoff value for good agreement is 6.0, indicated by the red line once again. Figures 15 and 16 show the average deviations about the mean and the median by individual severity.

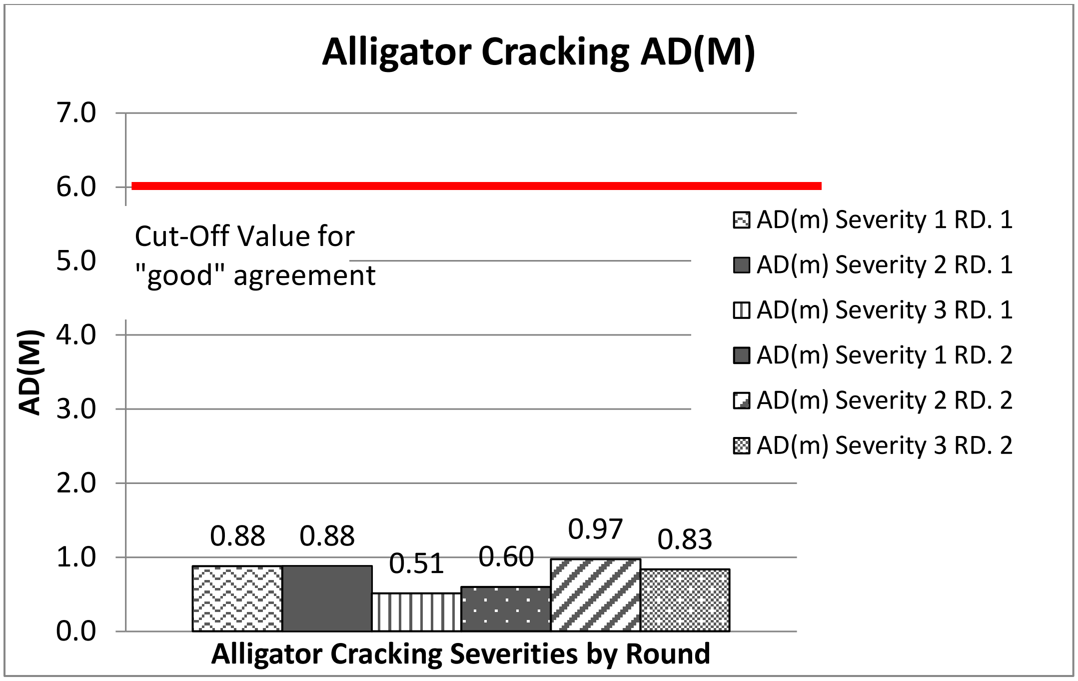


Figure 15: Alligator Cracking AD(M) by Severity and Round

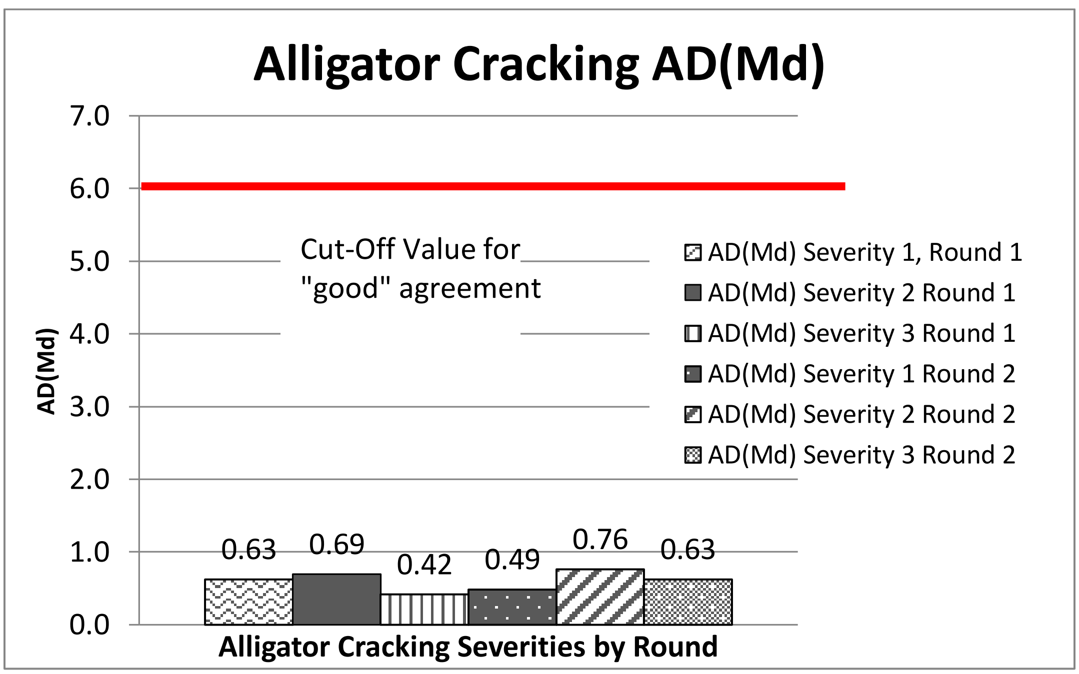


Figure 16: Alligator Cracking AD(Md) by Severity and Round

Large deviations in either of these graphs would indicate that the criteria between severities are unclear or that individual raters were having difficulty rating this distress in terms of the severity. The values are very close together, and very small, which shows that, in general, alligator cracking can be rated consistently and with little variation between raters.

The following figures, 17 and 18, show the average deviations about the mean and median for the total amount of alligator cracking found in each milepost plotted on the same graph. These values take severity out of the equation entirely and focus purely on the totals, giving another measure of variability within the protocol.

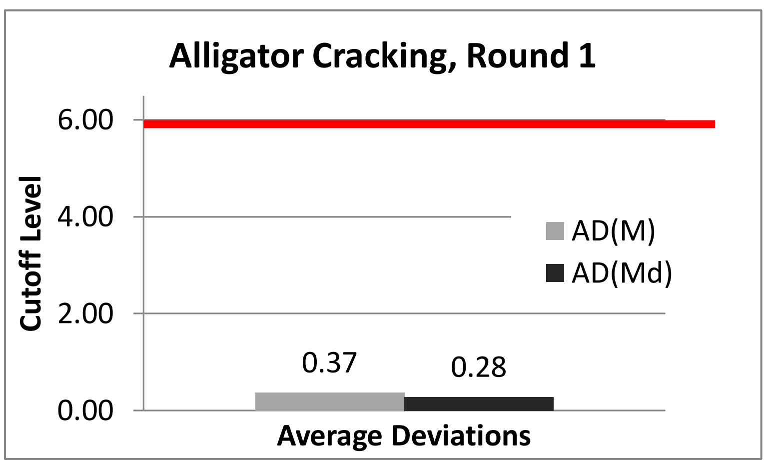


Figure 17: Total Alligator Cracking, Round 1

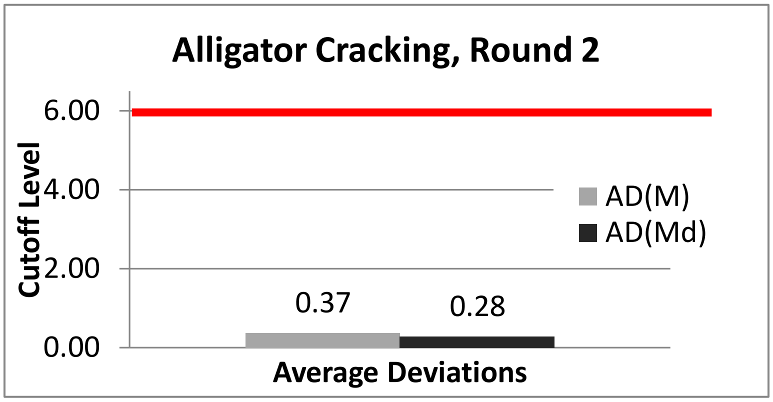


Figure 18: Total Alligator Cracking, Round 2

The average deviations for both rounds are exactly the same, even after multiple data checks, and determined to be occurring because of the nature of the data translation between total amounts of cracking found in the sites and the HPMS reporting requirements, which were the data items being tested. These deviations are slightly lower than the individual severity deviations, which may indicate that the problem is with rating the severity of the distress, not distinguishing quantity. A possible explanation of this is the location of asphalt distresses. A primary assumption in rating this type of distress is that it occurs only in the wheelpath, but distinguishing the wheelpath from the rest of the lane can be difficult depending on lane width, configuration (i.e. a curve in the road), and the type of facility being rated (busier roads with more vehicular traffic make evaluating the pavement more challenging). Because the location of the crack may be hard to identify from a roadside evaluation, it is possible that some alligator cracking was rated as longitudinal cracking or vice versa, since longitudinal cracking can appear identical to alligator cracking.

In order to compare the results of the new method to the old method, a simple arithmetic conversion is needed. Note that NMDOT wishes to combine longitudinal cracking that occurs in the wheelpath with fatigue cracking, which is consistent with the protocol for HPMS reporting. Pavement raters will have recorded lengths of this distress for each severity present in the test section. A single longitudinal crack in the wheelpath is considered Low severity fatigue cracking, and will be assigned a width of 2 feet for area calculations, as recommended by FHWA.

The alligator cracking calculation produces an area of the test section affected by the distress. Because the wheelpaths comprise only 33.3% of the entire lane (which would round up to 35%), a scale must be assigned in order to accurately obtain a value to be used for the NMDOT system, which operates on a 0 – 100% scale. A linear scale is being suggested at this point, with the conversions represented in Table 6.

Table 6: Alligator Cracking Extent Conversion

<table>
<tr><th>Percent Area – HPMS</th><th>Extent - NMDOT</th></tr>
<tr><td>0</td><td>0</td></tr>
<tr><td>5</td><td rowspan="2">1 (1-30 %)</td></tr>
<tr><td>10</td></tr>
<tr><td>15</td><td rowspan="2">2 (31-60%)</td></tr>
<tr><td>20</td></tr>
<tr><td>25</td><td rowspan="3">3 (61-100%)</td></tr>
<tr><td>30</td></tr>
<tr><td>35</td></tr>
</table>

The original values that the evaluators reported are converted to a percent area to be used for HPMS. The percentages are then converted into values that the NMDOT uses for its

PMS – the 0 through 3 system. An example is shown below, for clarification. Table 7 is a data sample taken from Round 1, already calculated into Percent Area for alligator cracking. The route is NM0041, sample section taken at milepost 32.

Table 7: Sample Data for Alligator Cracking

NM0041 – Milepost 32				
Severity	0	1	2	3
Square Feet	-	48	1320	40

The values will be manipulated mathematically to obtain a value used for HPMS reporting. The equation used to derive this number is below. The denominator = 6336 square feet for test sections of 1/10 mile with 12-foot lanes.

$$\text{Cracking Percent (CP)} = \frac{\sum(\#\, paces * (pace) * [\#WP * 2\ feet])}{Sample\ length * Lane\ width}$$

Note the "sum" symbol in the equation. This is to combine all severities reported to produce one value independent of severity for HPMS. NMDOT requires each severity be reported. Milepost 32 has 1320 square feet of Severity 2 alligator cracking. Conservatively, the number is rounded up to the next 5% increment. This is most advantageous when there is, say, 2.3% of a section affected. It will not get ignored by rounding to the nearest 5%, which would be 0.

$$\text{CP} = \frac{1320}{6336} = 20.8\%, \text{ round to } 25\%$$

By following Table 5: Alligator Cracking Extent Conversion, a 20% area result for HPMS use falls into NMDOT extent 2 (31-60%). The data item to be reported for NMDOT use would be [2, 3].

The same statistical analysis was performed on the translated values to determine whether the data translation process is valid. The cutoff value for agreement is once again 0.67 since the number of options is back to 4. The figure below shows the resulting AD(M) and AD(Md) for each severity after the reported values have been converted to NMDOT's scale. The vertical axis shows the acceptable cutoff point.

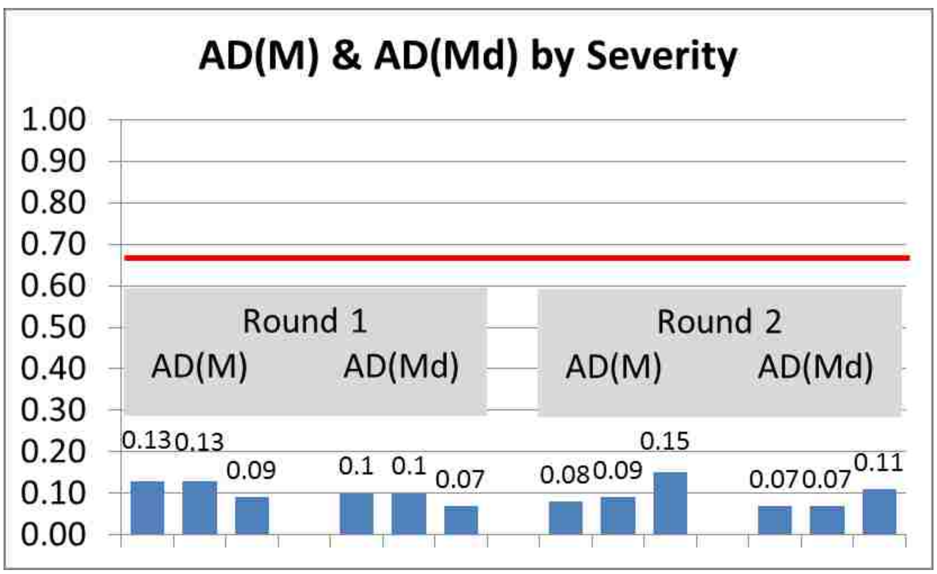


Figure 19: Translated AD(M) and AD(Md) for Alligator Cracking

After conversion to the NMDOT system, the variability among raters is completely acceptable, even after the rounding of the percent areas in the first step of the conversion that converts the areas to HPMS percentages. Since this measure is extremely important for Federal reporting, it is vital that this distress can be evaluated consistently. This also ensures that the state of New Mexico is obtaining extremely reliable data.

It is interesting to note that the variability is higher when severity levels are eliminated from the analysis entirely. Figure 20 shows the variability of total alligator ccracking throughout a section by rater.

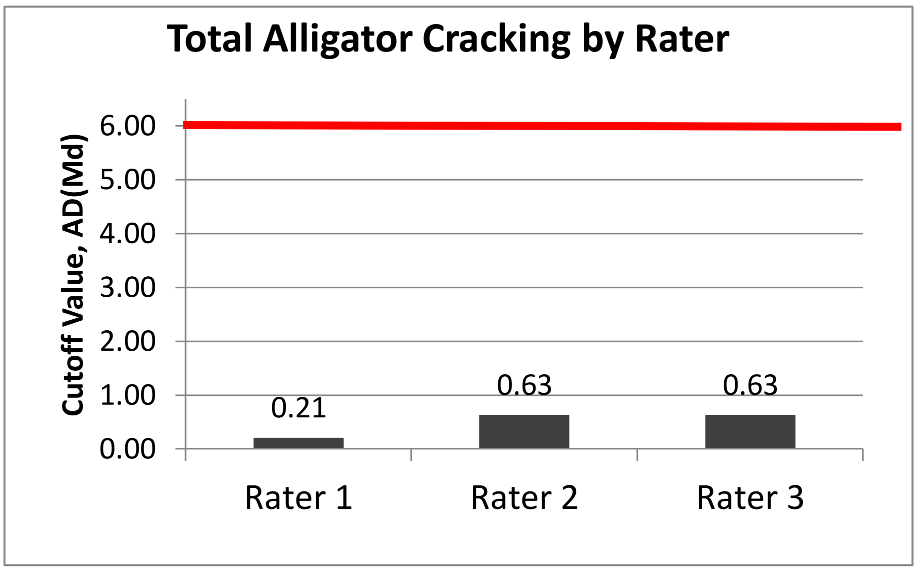


Figure 20: Total Alligator Cracking by Rater

It seems that the problem lies more within the actual identification of alligator cracking rather than determining its severity. This is primarily due to the changes in the protocol that classify longitudinal cracking within the wheelpath as severity 1 alligator cracking. On some sections, especially ones with large amounts of distress, it can be challenging to determine whether or not a longitudinal crack falls within the wheelpath or lies outside of it, and also how many linear feet of a particular crack lie within that wheelpath. However, the disagreement is so small that these findings may be insignificant in the big picture.

Transverse Cracking

Transverse cracking was analyzed in a similar manner as alligator cracking, with both a severity analysis and a total amount value. The following figures, 21 and 22, show the average deviations about the mean and median across two rounds of evaluations, by severity. Remember that the cutoff value for good agreement is 1020.0, indicated by the red line. The x-axis depicts the average deviations about the mean and median, and the y-axis shows the average deviations and cutoff levels for each analysis.

The values that were calculated fall well below the cut-off value, demonstrating that the agreement between raters is very good. It is also worthwhile to note that the agreement measure improves from the first to the second round of evaluation. This is an indicator that raters' evaluations improve as time passes. The vertical axis is AD(M).

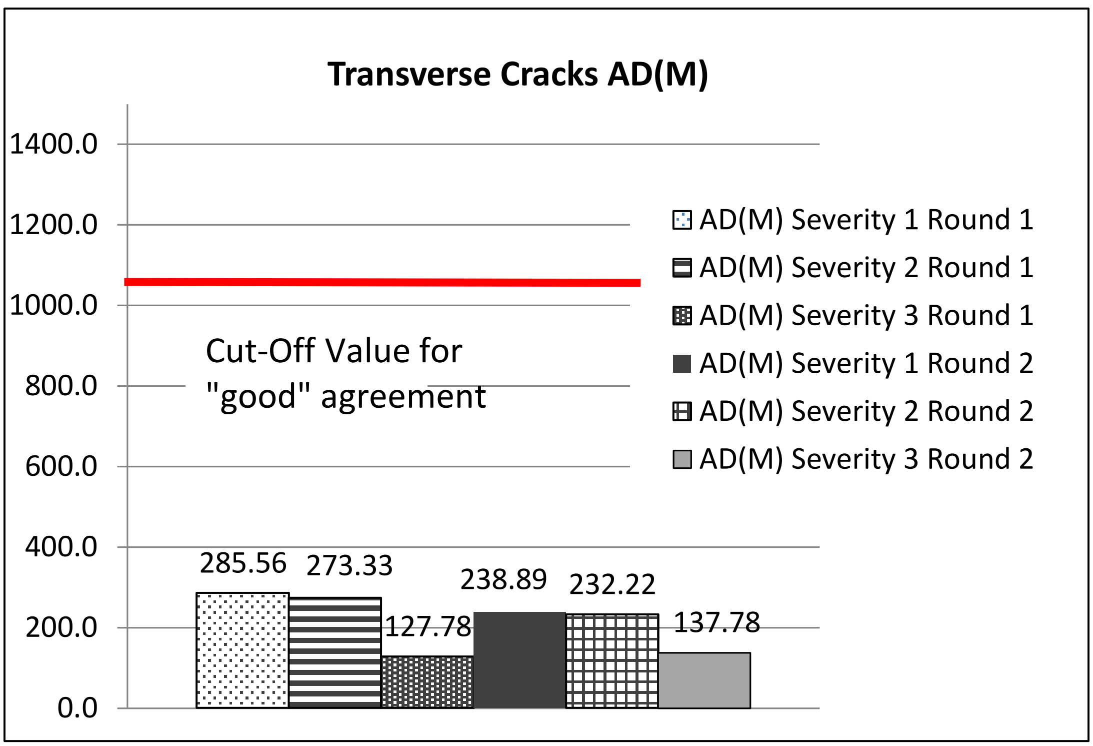


Figure 21: Transverse Cracking AD(M), Through 2 Rounds

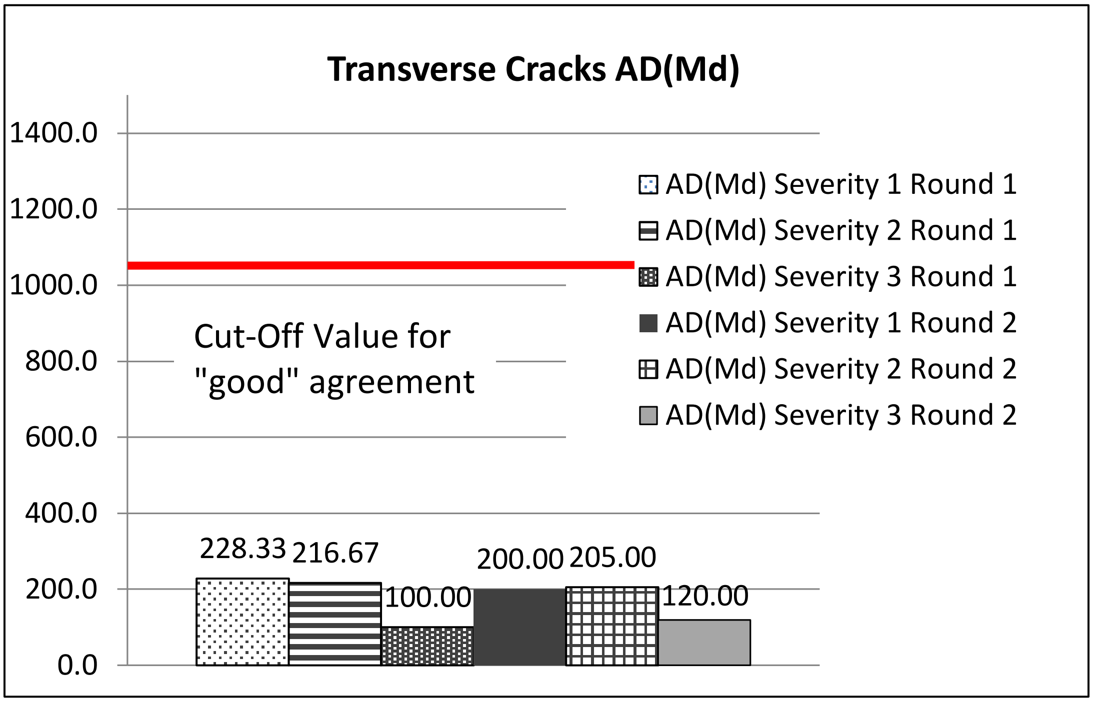


Figure 22: Transverse Cracking AD(Md), Through 2 Rounds

Figures 23 and 24 show the average deviations about the mean and median for transverse cracking as a whole, with severity completely ignored. This measure is to indicate whether the problem with rating transverse cracking falls within the severity or quantity domain. The y-axis for both graphs depicts the average deviations and the cutoff levels.

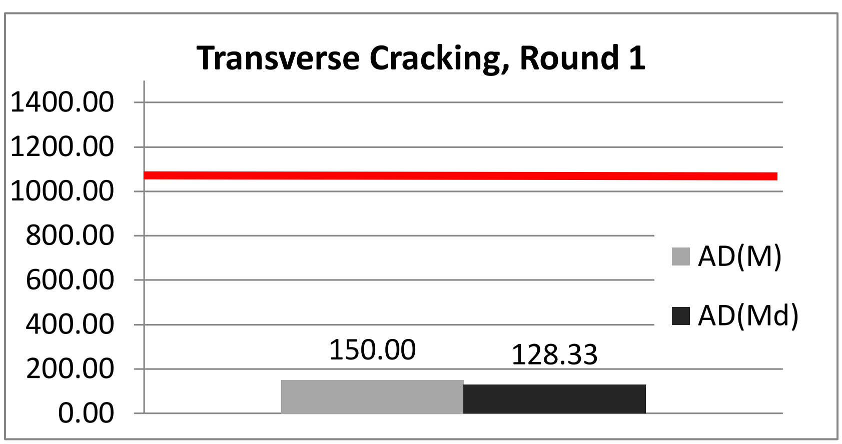


Figure 23: Total Transverse Cracking Round 1

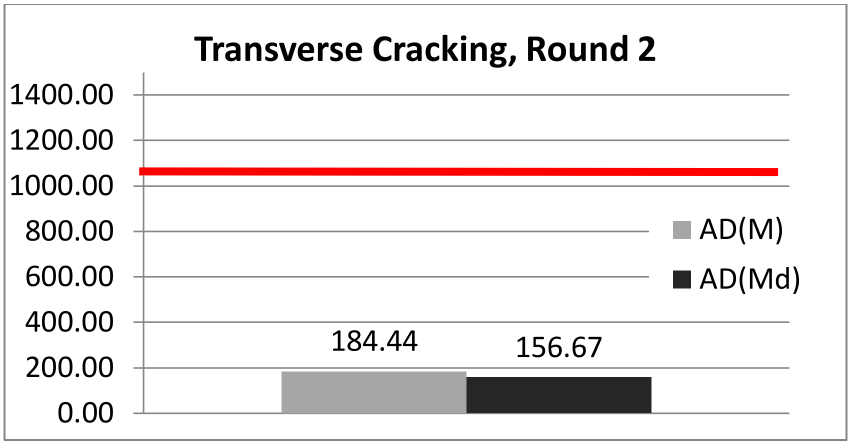


Figure 24: Total Transverse Cracking Round 2

The average deviations for both rounds are slightly higher than the individual severity deviations, which may indicate that the problem is not with rating the quantity of the distress, but distinguishing between the severities. A possible explanation of this discrepancy may be that it is up to the evaluator to decide which severity the crack falls into, while trying to keep up speed and peer across an entire lane with traffic interrupting the view. New to this protocol is also the "10% Rule," which states that if 10% of a crack can be rated as a higher severity, then it shall be rated as the higher severity. Another possibility of variability in general may be that the evaluator must decide whether or not a crack reaches halfway across the lane in order to be counted, according to HPMS guidelines. This seemingly simple judgment call is made much more challenging when a section of road is in poor condition and it is almost impossible to determine where one crack ends and another begins.

The second part of analyzing this distress is the conversion of the data to a form that works in NMDOT's pavement management system. Since NMDOT wants percent of test section area affected by this type of distress rather than linear feet of cracking, a base assumption had to be made. In order to estimate the maximum number of cracks that may occur within a test section, Nevada's Pavement Evaluation Manual was referenced. Nevada assumes that severe transverse cracking is present if a transverse crack occurs every 5 feet. If the most severe case of transverse cracking is happening, a transverse crack occurs every 5 feet, therefore there will be up to 106 cracks present in the section. This methodology was used to create cutoff values according to the number of cracks located in a section, however, the actual numbers were adjusted based on previously recorded data from previous years. Using percentages of test section area that correspond with the NMDOT's extent reporting values, and the number of cracks that have previously coded as each extent, we can set guidelines for how many cracks fall into each extent. It was discovered that, on average, less than 8 cracks were rated as an Extent 1, 9 to 16 cracks were reported as Extent 2, and 17 or higher cracks within a section warranted a Severity 3 rating. These threshold values for transverse cracks were determined from the data of visual surveys: severity and extent ratings based on the old protocol, and severity and crack counts from the proposed (new) protocol. The data from 6 raters obtained in round 2 of the distress surveys (NMSU and UNM raters) were used in this part of the analysis. For each section, the mean value of the crack count (new protocol) with highest severity present was compared to the extent rating (old protocol). For the data available, the mean values of crack count were 6.5, 14.0 and 19.0 for extent ratings of 1, 2 and 3 respectively. The standard deviations were 1.7, 2.4 and 1.7 for extent ratings of 1, 2 and 3 respectively. For severity 1, the mean plus standard deviation was 8.2 cracks (approximated to 8), and the mean minus standard deviation for severity 3 was 17.3 (approximated to 17 cracks). The results are listed in Table 8.

Table 8: Extent Ranges for Transverse Cracking

NMDOT Extent	Minimum # of Cracks	Maximum # of Cracks
0	0	0
1	1	8
2	9	16
3	17	(none)

The number of cracks of a particular severity can be sorted into the appropriate Extent for reporting. Extent 0 indicates the distress is not present.

The original values that the evaluators reported are converted to a linear footage to be used for HPMS. These numbers are then converted into values that the NMDOT uses for its PMS – the 0 through 3 system. An example is shown below, for clarification. Table 9 is a data sample taken from Round 1, number of cracks reported of longitudinal cracking. The route is NM0041, sample section taken at milepost 3.

Table 9: Sample Data for Transverse Cracking

NM0041 – Milepost 3				
Severity	0	1	2	3
# Cracks	-	38	2	0

The pavement raters will simply record the number of each severity of crack present on the test section. This can be easily manipulated to produce the data value that HPMS requires. A simple algorithm can easily translate the data into the NMDOT format as well. By simply following Table 6: Extent Ranges for Transverse Cracking, 3 data pairs will be created: [1, 3], [2, 1], and [3, 0].

After the evaluators have reported the number of cracks, the new extents are translated into NMDOT's rating scale. Figure 25 shows the results of this analysis. Once again, the y-axis depicts average deviation and the cutoff level for agreement.

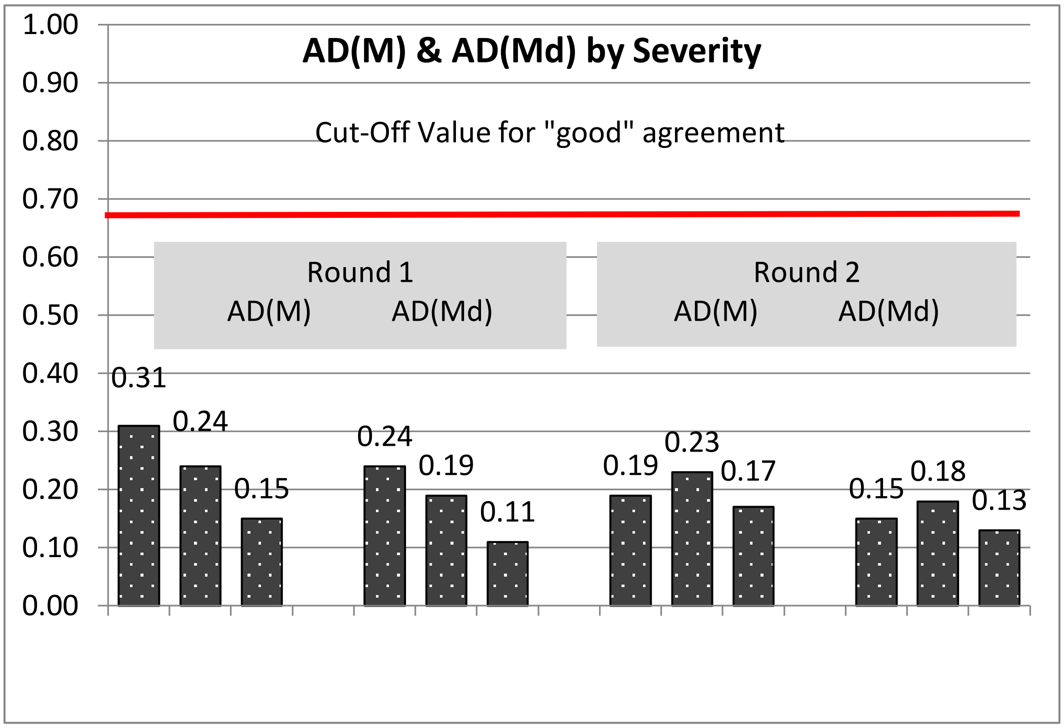


Figure 25: Translated Transverse Cracking AD(M) and AD(Md) By Round

All of the translated values fall well within the accepted variability threshold of 0.67.

Raveling/Weathering

Raveling and weathering is a distress that did not change much from the old to the new protocol. The rating criteria were revised so that the severity of this distress could be rated with more accuracy, and a new provision was allowed that indicated none was present, as the old protocol stated that all roadways, regardless of age or appearance, had to be rated as a severity 1. As a result of the application of the new rating criteria, variability in the severity aspect was reduced among raters, as shown in Figure 26.

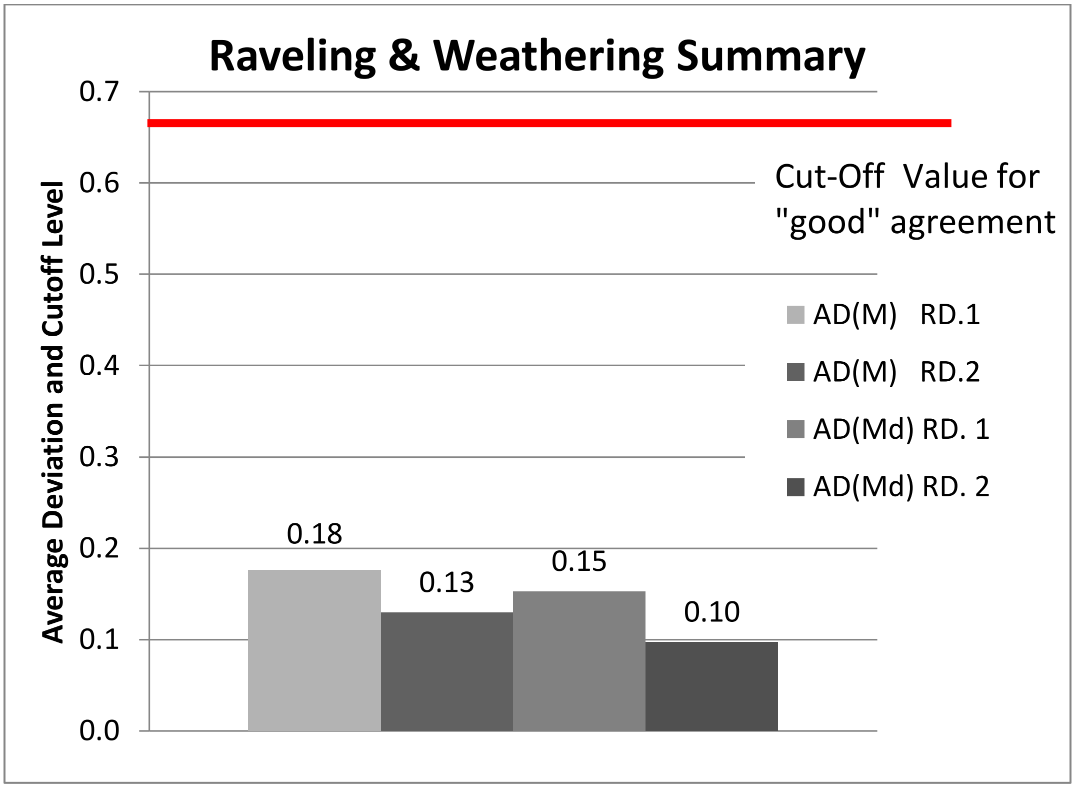


Figure 26: Raveling & Weathering AD(M) and AD(Md)

Since the assumed extent has been changed to the entire section, there is now zero variability in this characteristic of the pavement rating.

Bleeding

Bleeding is still the hardest distress to quantify, due to the nature of the defect. In the new protocol, raters are to identify all severities that they encounter across the test section. This can be notoriously difficult. One rater may perceive a discoloration on the roadway as a spill, while another may perceive it as a Severity 3 – where no aggregate shows through. An evaluator may miss an occurrence entirely, while one may catch them all.

Even more complexity is added when raters are asked to note all three severities rather than the most severe. However, because of revised rating criteria, the variability seen previously during the rating of this distress has been reduced. Figure 27 shows the most recent analysis of interrater variability.

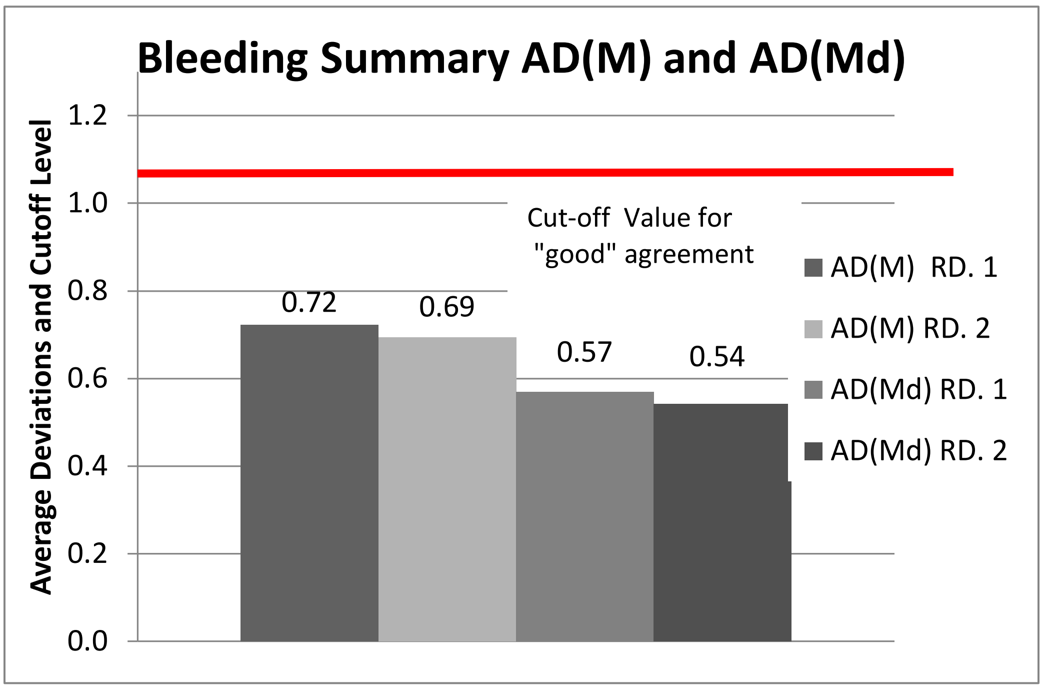


Figure 27: Bleeding AD(M) and AD(Md) By Round

The figure above only depicts the variability among severities, as extent is assumed to be Extent 1 in all cases, as explained by the analysis of previous data, and therefore reduces variability to zero.

Longitudinal Cracking

The rating procedure for longitudinal cracking did not change, except that now this distress is to include edge cracking, which is essentially longitudinal cracking that occurs within one foot of the edge stripe. Now that data for longitudinal cracking requires that each severity be evaluated, the assessing of interrater agreement becomes more problematic. It is more difficult for the raters to keep track of each severity and extent that occurs, therefore the variability associated with rating this distress was expected to increase. Couple that with the rater's ability to distinguish the exact location of a wheelpath, and you will get mixed results.

This issue is supported by the results shown in Figure 28 for severity ratings. It can be seen that the agreement among raters increased significantly and dropped below cut-off value (c = 3.17) in round 2 (shown on the y-axis). This is likely due to the experience and familiarity about the criteria and procedures gained by the raters with time and training.

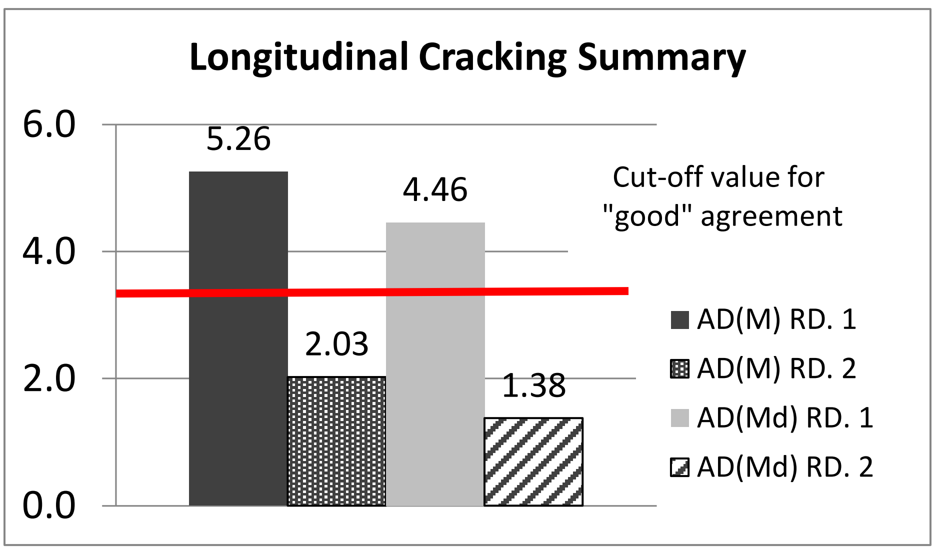


Figure 28: Longitudinal Cracking, Rounds 1 and 2

4.1.3 Interrater Agreement Comparison Between Protocols

Four of the distresses rated using the new protocol can be directly translated back into the old protocol based on numerical analyses and simple mathematics. In order to judge the success of the new protocol, the average deviations about the median were compared and shown in Figure 29.

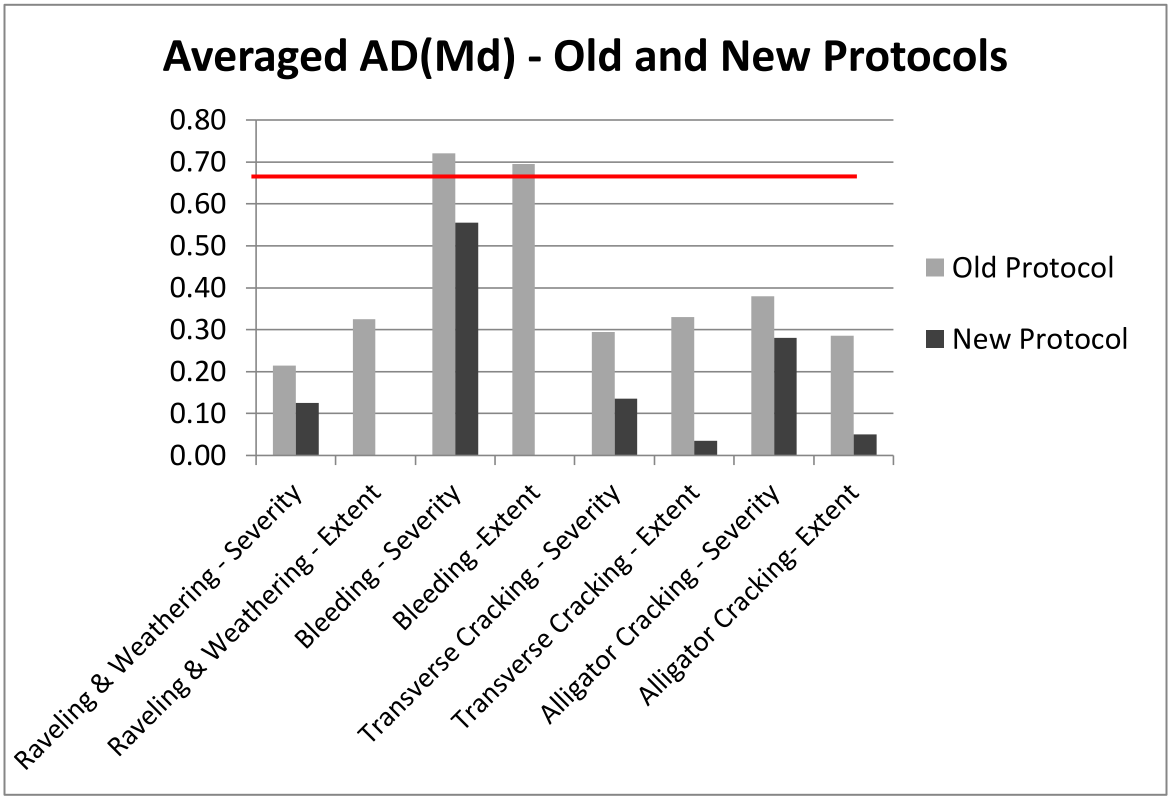


Figure 29: Averaged AD(Md) – Old and New Protocols

With every distress tested, the variability decreased when the new protocol was applied. The variability for Raveling & Weathering – Extent and Bleeding – Extent was eliminated due to the analysis performed on the 2006-2009 data that allowed assumptions to be made regarding those values.

4.2 Agreement Over Time

Previously, results were obtained by comparing raters to each other. In this analysis, each rater's evaluations were compared by round in order to determine whether the perspective of the rater changes over time. For example, Rater 1 should be able to consistently evaluate alligator cracking, severity 1 regardless of when he performs the evaluation. The following paragraphs will demonstrate the results of this analysis by individual distress. It is important to note that the same testing methodology will be used, the average deviation indices, and that a smaller number indicates that the individual rater is more consistent. The data used in the following analyses are the ratings obtained from the new protocol, rounds 1 and 2, for each evaluator. Because there are only 2 data points, the mean and the median for each set are the same, which makes the average deviation about the mean and median also the same, which is why only one set will be shown.

Alligator Cracking

The amount (total area) of alligator cracking recorded by each rater by test section was compared between the two rounds of data collection and separated by severity. The average deviations are shown in Figure 30 with the average deviation and cutoff level depicted on the y-axis.

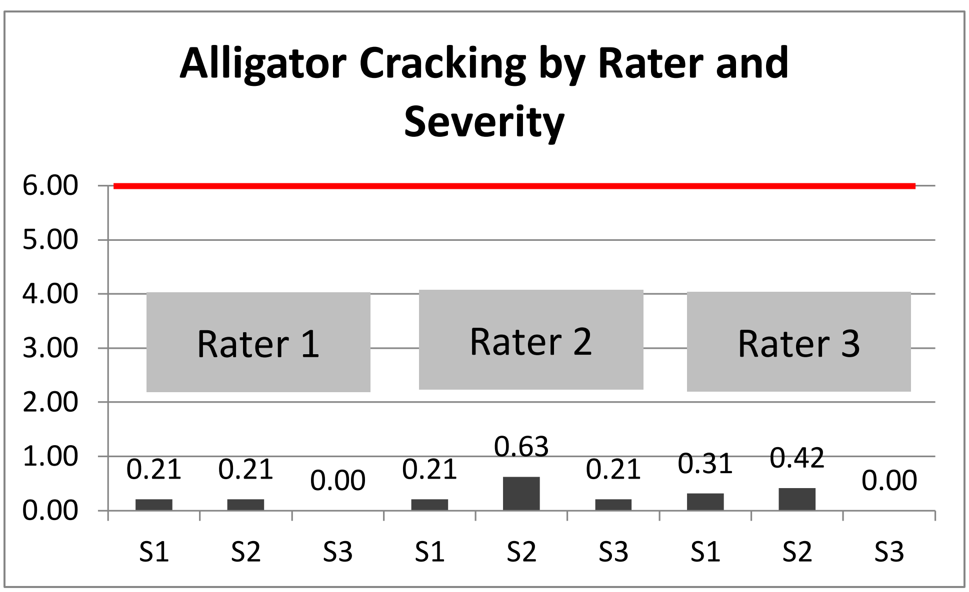


Figure 30: Alligator Cracking by Rater and Severity

All three raters' observations fall well below the acceptable level of variation. From this graph, it can be deduced that severity 2 is the most variable from round to round. When the total amount of alligator cracking within a section is compared between rounds, the variability is slightly different, shown in Figure 30.

All three raters can accurately rate the same section twice, as the average deviations are within the acceptable limits. It is believed that the deviations are higher than the group deviations because there are fewer data points to compare to; the raters only had 2 rounds. More rounds of rating the same section would most likely reduce the standard deviation. See Figure 31.

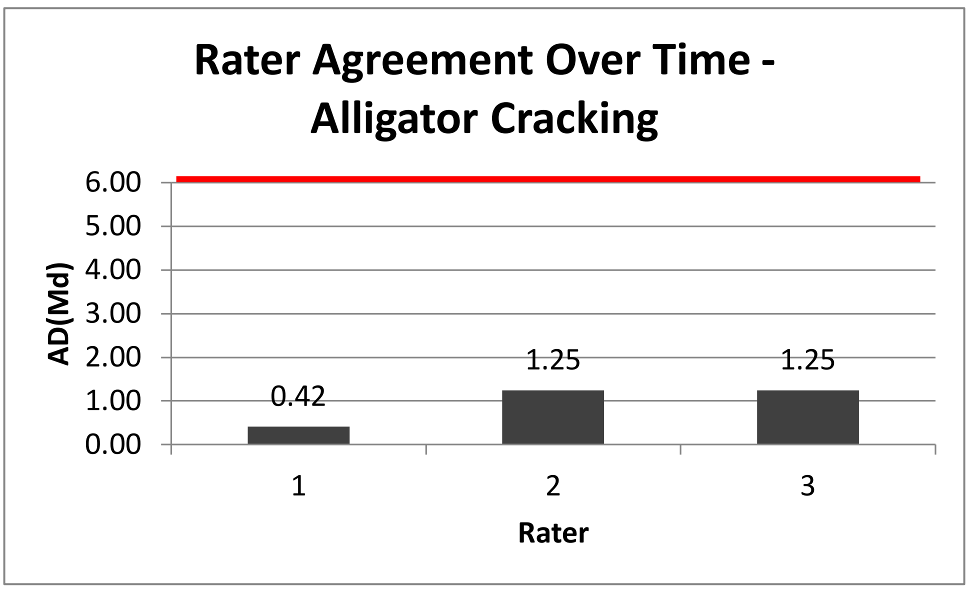


Figure 31: Rater Agreement over Time – Alligator Cracking

Transverse Cracking

This distress was evaluated using the number of cracks that each rater reported in each test section. The objective of this analysis was to see whether or not raters could identify the same number of cracks in the same section by severity consistently. The results of the average deviation analysis are shown below in Figure 32.

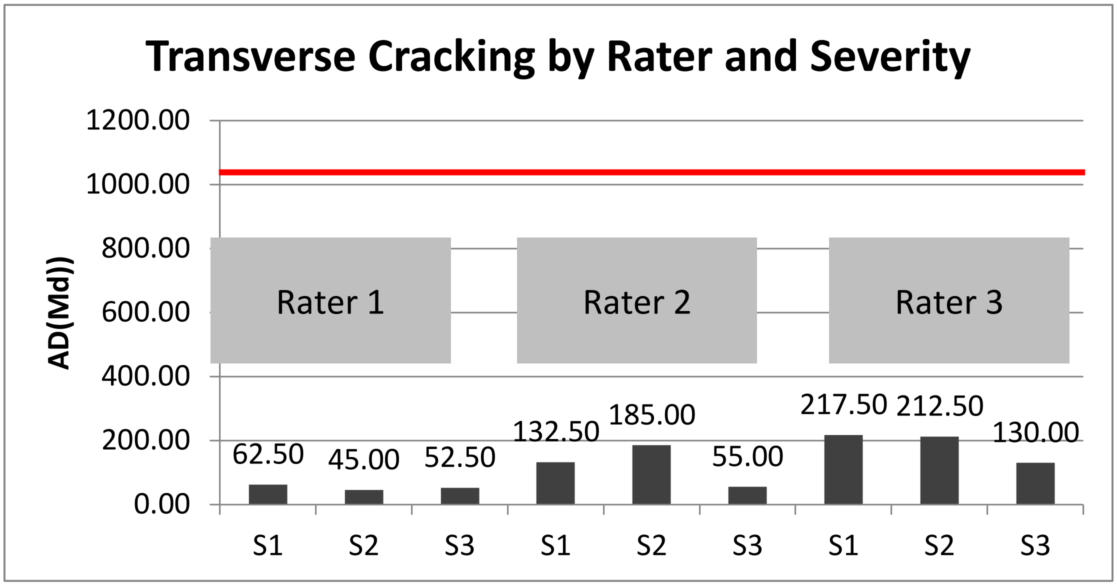


Figure 32: Transverse Cracking by Rater and Severity

While all of the measures fall well below the acceptable cut-off value for transverse cracking, two out of three raters had more difficulty differentiating between severity 1 and severity 2 cracks. Upon closer review of the data, the severities with the highest variability between rounds had counts that had reversed from one round to the next. For example, Rater 2 had recorded 155 cracks in severity 1 and 166 cracks in severity 2 in the first round. The second round showed 78 severity 1 cracks and 116 severity 2 cracks. In order to determine whether the problem was in defining the severity of the cracks or whether or not a crack should be counted, the same analysis was performed using total number of cracks. The results are presented in Figure 33.

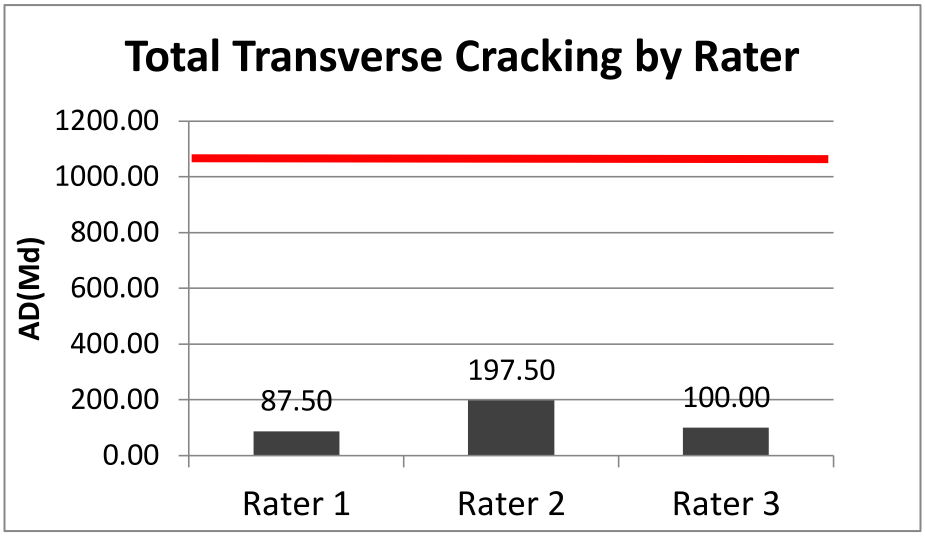


Figure 33: Total Transverse Cracking by Rater

The results are mixed – indicating that raters have trouble in certain areas. Raters 1 and 2 showed issues in determining whether or not to count half cracks, and Rater 3 had more trouble discerning between severities of cracks. Again, the variability measure falls well below the acceptable threshold, so the problem is far from causing issues with data quality, but perhaps more training should be done in the future to further decrease these inconsistencies.

Additionally, each rater was analyzed by his performance over time to see if the deviations were acceptable. The result of this analysis is shown in Figure 34.

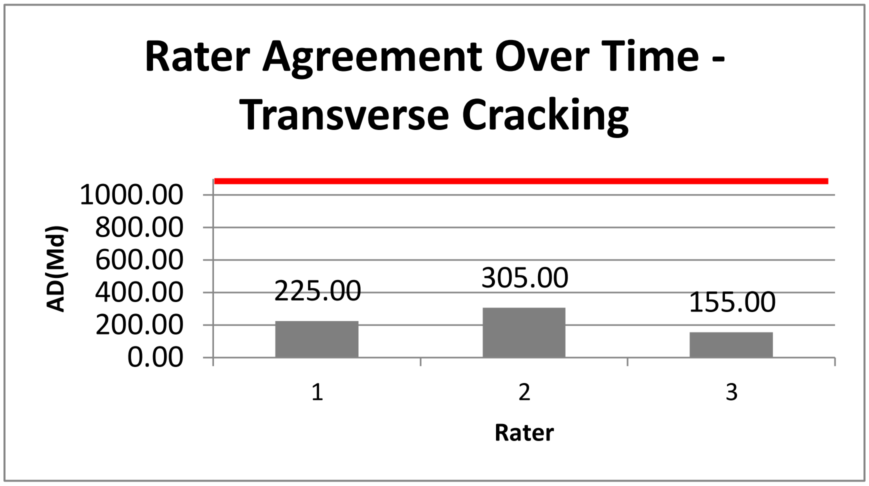


Figure 34: Rater Agreement Over Time – Transverse Cracking

The results show that the raters were able to rate the same sections relatively accurately, with the deviations falling well within accepted levels.

Raveling/Weathering

The results for the agreement between raters regarding raveling and weathering are shown in Figure 35. The results for this distress are mixed, depending on the rater. Rater 1 has smaller variability than the group as a whole, Rater 2 is on par with the group, and Rater 3 exceeds the group's average inconsistency. The values are still well within the threshold for acceptable data quality.

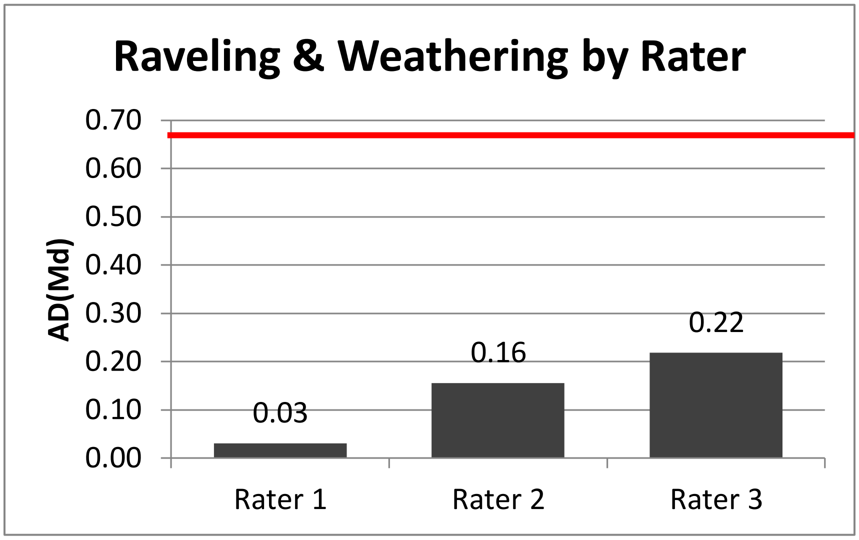


Figure 35: Total Raveling & Weathering by Rater

Raveling and weathering was also checked to determine if raters could consistently evaluate this distress over time. The results are shown in Figure 36 below:

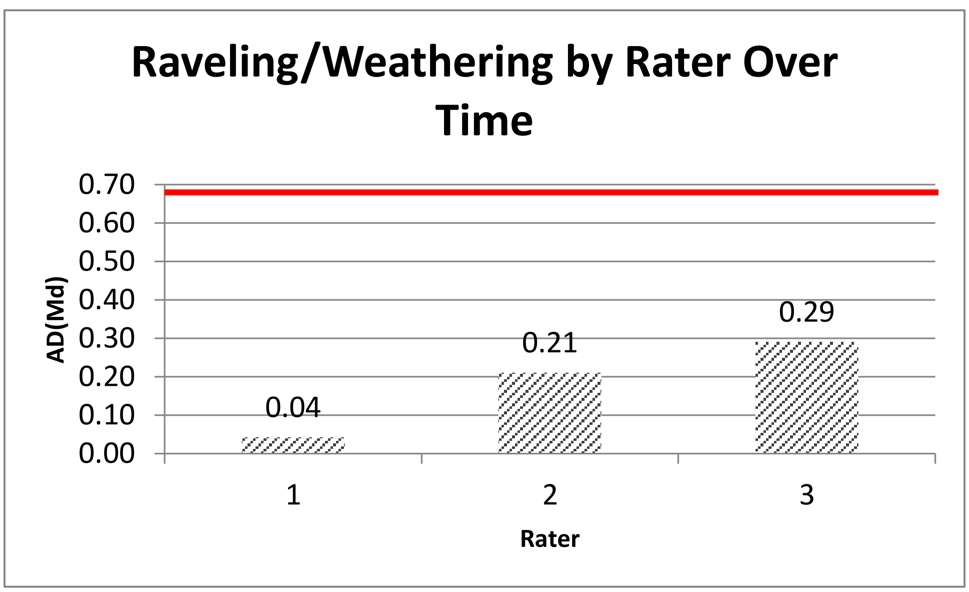


Figure 36: Raveling/Weathering by Rater over Time

As the figure shows, all raters are consistently able to evaluate raveling and weathering over time. Raters two and three have more trouble being consistent than the overall group results, which may be the rater or the fact that there are only two rounds to compare to, rather than several raters over several rounds being compared at once.

Bleeding

This distress showed a greater variability than the rest, not surprisingly. The raters continued to have a difficult time rating bleeding consistently. Figure 37 shows the results. Since bleeding has more options and various numbers are assigned to each rating combination, there is a difference between the mean and median, therefore the average deviations about the mean and median are shown for completeness.

Because the deviations within rounds according to rater are higher than those of the interrater agreement, it seems the concern is with repeatability of the evaluation. Upon closer inspection of the data, the raters are not consistently identifying and/or recording all instances of bleeding that occur within the test section. This could be due to traffic, time of day, or other distractions that occur while rating pavement from the roadside.

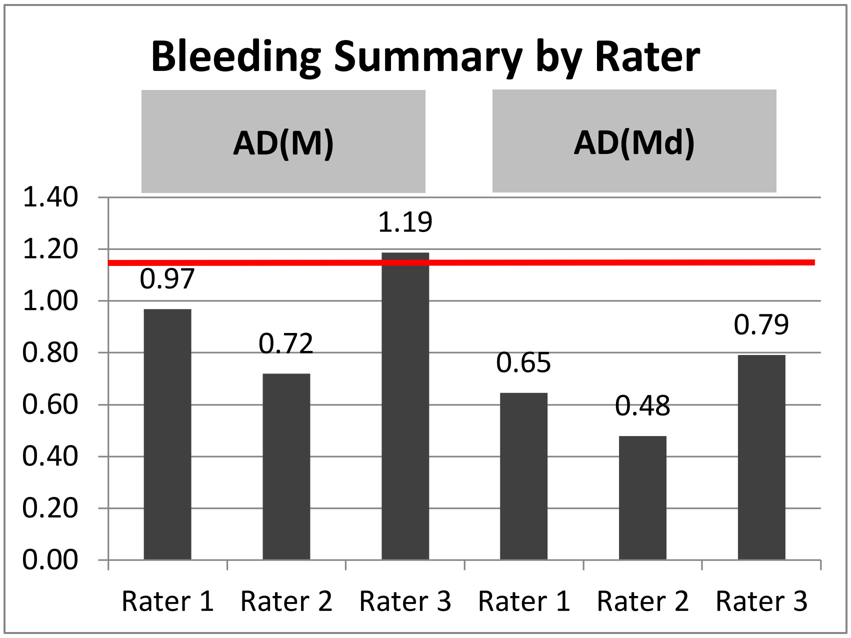


Figure 37: Bleeding Summary by Rater

Agreement over time was evaluated for this distress also. The results are shown below, in Figure 38.

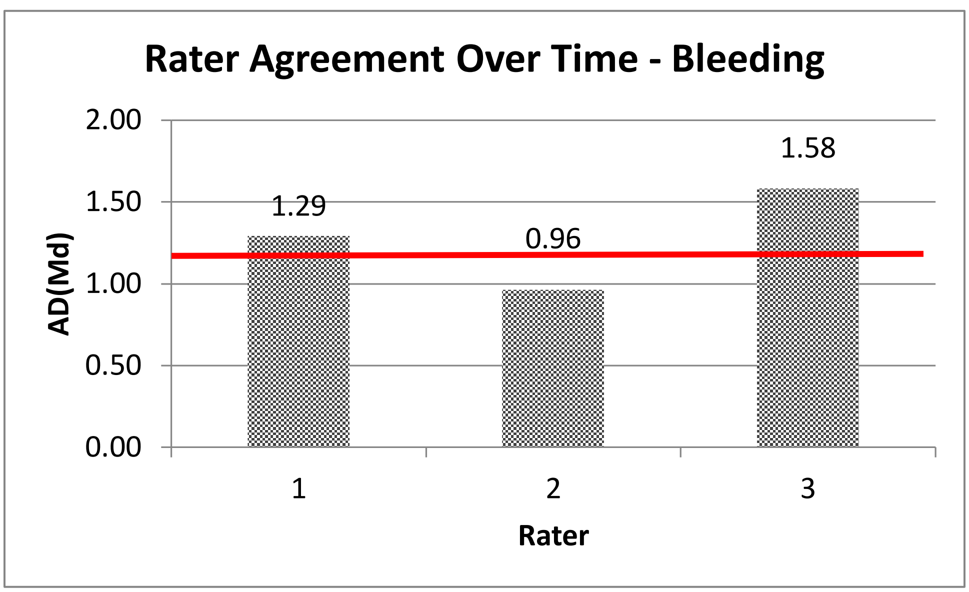


Figure 38: Rater Agreement Over Time - Bleeding

Again, this evaluation follows the trend of rater agreement over time having a higher average deviation than the whole group. This analysis shows that only one rater has acceptable results. The outcome of this is also not surprising, considering the level of difficulty in rating this distress to begin with, and, with more time, the average deviations are expected to go down.

Longitudinal Cracking

The agreement analysis for longitudinal cracking fell in line with the rest of the results. For the most part, the raters were consistent, and the values are within tolerances. See Figure 39 for details.

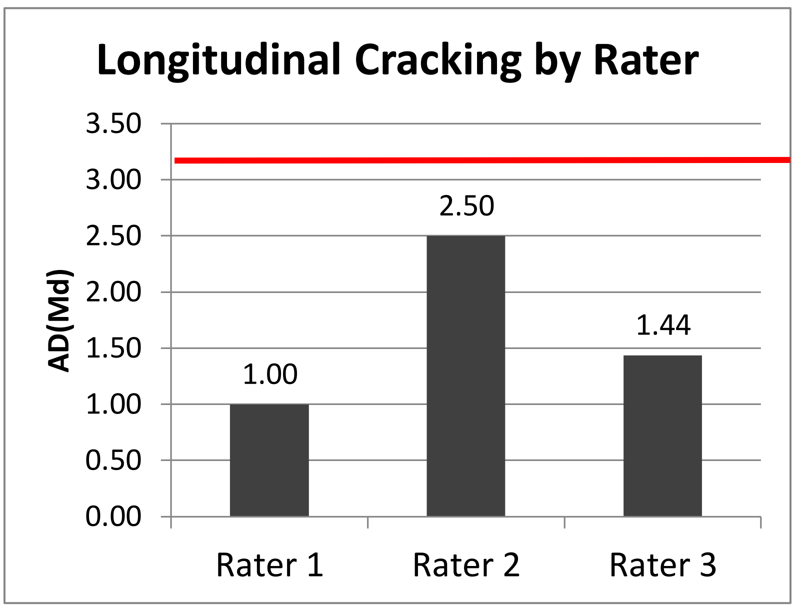


Figure 39: Longitudinal Cracking by Rater

The majority of the average deviations for one particular rater across time are smaller than those of the interrater analyses, regardless of distress. This would suggest that the disagreement lies primarily between raters and not indicative of changing perceptions by any particular rater. This end result demonstrates that the newly written criteria are more robust than the previous set, less objective, per se, and that there is room for improvement in group calibration activities.

The same analysis was performed to confirm rater agreement over time wth longitudinal cracking. The results are shown in Figure 40. The results are mixed; two of the three raters are able to evaluate this distress within acceptable limits. The main cause for this disagreement is most likely the changes in protocols, as now longitudinal cracking is combined with edge cracking, and the rater may be unsure of the location of the cracking, marking it as wheelpath rather than outside the wheelpath.

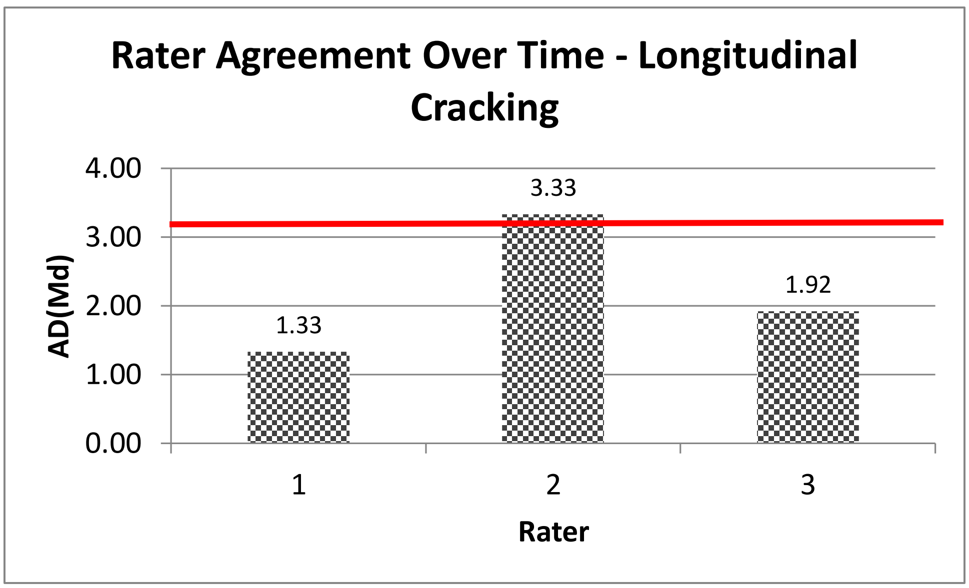


Figure 40: Rater Agreement over Time – Longitudinal Cracking

4.3 Rating Estimation vs. Actual Measurements

4.3.1 Transverse Cracking Analysis

The new protocol requires that evaluators count only cracks that are half lane width (6 ft) or more, leaving the shorter cracks out of the rating. This brought up the concern that perhaps the overall amount of cracking would be under- or over-represented by making this decision. On one hand, there may be plenty of smaller cracks that make up for this difference in cumulative length; on the other, the evaluation could be grossly overestimating total length per mile if there are no smaller cracks to make up the discrepancy. An additional concern was that regarding the shape of the cracks. Typically, transverse cracks are not straight-line cracks; they are irregular and they tend to have a sinuous, wavy shape. An extra set of data was collected to ensure this did not have a significant effect on the total crack length as well. Both of these factors were taken into consideration when determining if a correction factor was needed to calculate HPMS data from the estimated quantities of distress found throughout the network.

The field measurements provided the actual length of transverse cracking per section for this analysis. Transverse cracks were measured and separated into 2 categories: cracks 6 feet and longer, and those under 6 feet. These measurements are sorted by crack severity in Table 10.

Table 10: Field Measurements and Estimates of Transverse Cracking Length

Route	Milepost	Severity	Estimated Length by Raters	Actual Length, >6ft	Actual Length, All	Difference, Estimate and >6ft	Difference, Estimate and All
NM 28	17.0	1	192	229.6	305.1	37.6	113.1
NM 28	17.1	1	412	374.8	491.2	-37.2	79.2
NM 28	17.2	1	356	384.9	466.9	28.9	110.9
I 25	15.0	1	80	109.5	134.4	29.5	54.4
I 25	15.1	1	100	73.3	146.1	-26.7	46.1
I 25	15.2	1	48	37.8	67.0	-10.2	19.0
I 10	129.0	1	88	70.7	92.8	-17.3	4.8
I 10	129.1	1	24	25.1	60.8	1.1	36.8
I 10	129.2	1	48	32.1	67.5	-15.9	19.5
I 10	129.3	1	44	22.9	54.0	-21.1	10.0
US 70	144.8	1	32	16.5	164.8	-15.5	132.8
US 70	144.9	1	64	44.9	346.7	-19.1	282.7
NM 478	19.0	1	276	293.8	305.8	17.8	29.8
NM 478	19.1	1	342	339.0	375.7	-3.0	33.7
NM 478	19.2	1	360	375.3	408.0	15.3	48.0
NM 28	17.0	2	28	39.8	65.8	11.8	37.8
NM 28	17.1	2	60	119.0	143.8	59.0	83.8
NM 28	17.2	2	8	6.0	6.0	-2.0	-2.0
I 25	15.0	2	0	0.0	0.0	0.0	0.0
I 25	15.1	2	0	0.0	0.0	0.0	0.0
I 25	15.2	2	0	0.0	0.0	0.0	0.0
I 10	129.0	2	0	0.0	0.0	0.0	0.0
I 10	129.1	2	8	0.0	0.0	-8.0	-8.0
I 10	129.2	2	0	0.0	0.0	0.0	0.0
I 10	129.3	2	0	0.0	0.0	0.0	0.0
US 70	144.8	2	40	36.3	55.7	-3.7	15.7
US 70	144.9	2	84	95.5	126.3	11.5	42.3
NM 478	19.0	2	54	133.5	133.5	79.5	79.5
NM 478	19.1	2	24	30.8	39.3	6.8	15.3
NM 478	19.2	2	0	0.0	0.0	0.0	0.0
NM 28	17.0	3	0	0.0	0.0	0.0	0.0
NM 28	17.1	3	0	0.0	0.0	0.0	0.0
NM 28	17.2	3	0	0.0	0.0	0.0	0.0

Route	Milepost	Severity	Estimated Length by Raters	Actual Length, >6ft	Actual Length, All	Difference, Estimate and >6ft	Difference, Estimate and All
I 25	15.0	3	0	0.0	0.0	0.0	0.0
I 25	15.1	3	0	0.0	0.0	0.0	0.0
I 25	15.2	3	0	0.0	0.0	0.0	0.0
I 10	129.0	3	0	0.0	0.0	0.0	0.0
I 10	129.1	3	0	0.0	0.0	0.0	0.0
I 10	129.2	3	0	0.0	0.0	0.0	0.0
I 10	129.3	3	0	0.0	0.0	0.0	0.0
US 70	144.8	3	16	0.0	5.7	-16.0	-10.3
US 70	144.9	3	28	9.4	9.4	-18.6	-18.6
NM 478	19.0	3	0	0.0	0.0	0.0	0.0
NM 478	19.1	3	0	0.0	0.0	0.0	0.0
NM 478	19.2	3	0	0.0	0.0	0.0	0.0

Table 10: Field Measurements and Estimates of Transverse Cracking Length, Continued

Figures 41 and 42 show the correlations between the estimated data obtained from the evaluators and the actual detailed measurements taken in the field. The estimates provided by the raters tend to underestimate the total length of all cracks in a section by around 25%, with a coefficient of determination (R^2) of 0.87. When the estimates are compared to the total length of cracks only 6 feet or longer, the correlation improves dramatically and the $R^2 = 0.97$. This makes sense because the parameters are the same in this analysis, and gives credibility to the raters' ability to accurately judge the amount of cracking within a test section using the new protocol.

The average cumulative difference between the raters' count and the detailed measurements was 2 feet and the standard deviation was 20.3 ft. Effectively, this amounts to only 2 cracks at the most that are responsible for the variance, which is very small when considering a 0.1 mile section. Finally, the length of transverse cracking for HPMS reporting can be estimated from the cumulative length estimated from raters' data using a multiplication factor of 1.25 (equation shown in Figure 36).

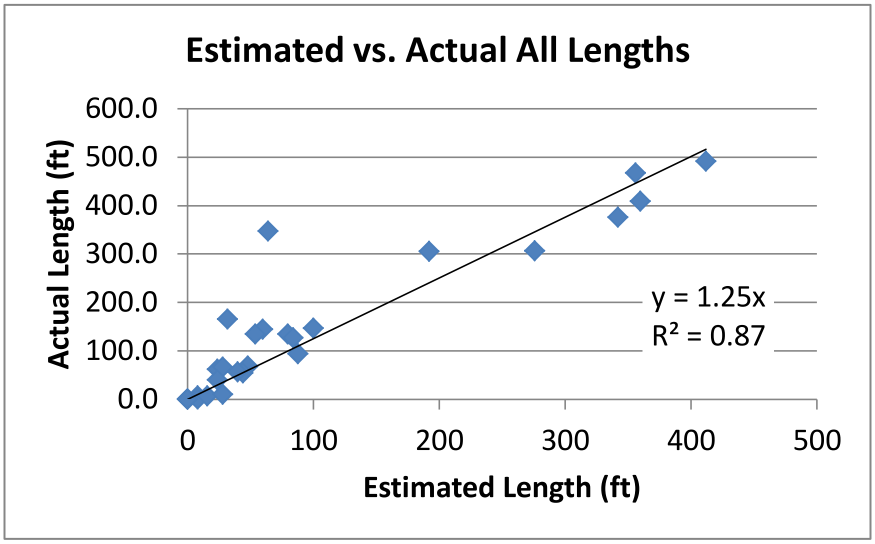


Figure 41: Raters' Estimated vs. Actual Length, All Transverse Cracking

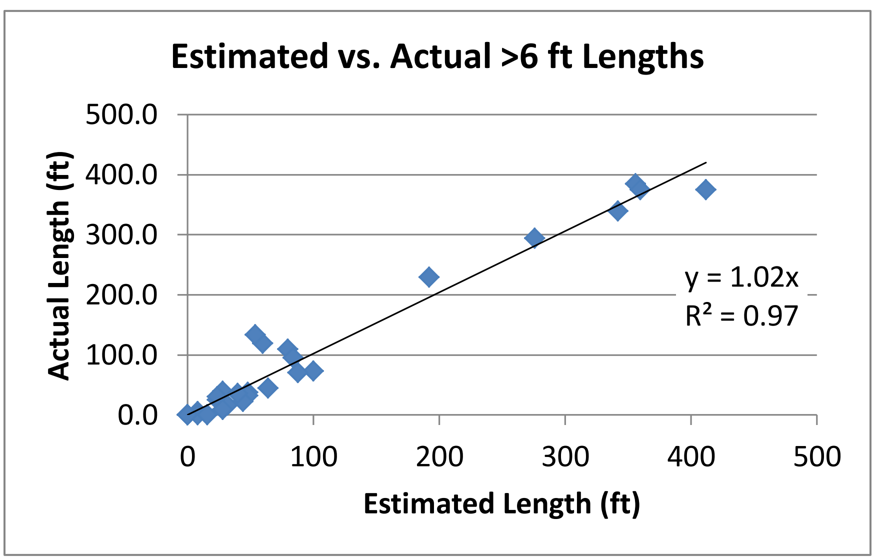


Figure 42: Raters' Estimated vs. Actual Length, Transverse Cracking >6 feet

4.3.2 Effect of Crack Geometry

The field measurements regarding 2-point (straight-line) versus actual crack length are given in Table 11. Figure 43 compares these lengths in order to determine whether the shape of the crack affects its length significantly enough to require an adjustment factor. In this figure, the segmented line represents the equality (both parameters are equal, y = x) and the continuous line is the best-fit line. The R^2 value is 1.00 indicating that the best-fit equation predicts the relationship between actual and 2-point length measurements for transverse cracks.

Table 11: Transverse Cracking 2-Point and Actual Length Field Measurements

Milepost	Severity	Estimated Length by Raters	Actual Length, All	2-Point Length	Difference	Percent Difference
17.0	1	192	305.1	293.9	11.2	3.7
17.1	1	412	491.2	462.2	29.0	5.9
17.2	1	356	466.9	433.4	33.5	7.2
15.0	1	80	134.4	124.6	9.8	7.3
15.1	1	100	146.1	134.7	11.4	7.8
15.2	1	48	67.0	60.9	6.1	9.1
129.0	1	88	92.8	88.7	4.1	4.4
129.1	1	24	60.8	56.3	4.5	7.4
129.2	1	48	67.5	63.6	3.9	5.8
129.3	1	44	54.0	49.4	4.6	8.5
144.8	1	32	164.8	153.9	10.9	6.6
144.9	1	64	346.7	324.6	22.1	6.4
19.0	1	276	305.8	289.1	16.7	5.5
19.1	1	342	375.7	360.0	15.7	4.2
19.2	1	360	408.0	378.7	29.3	7.2
17.0	2	28	65.8	62.9	2.9	4.4
17.1	2	60	143.8	136.6	7.2	5.0
17.2	2	8	6.0	5.9	0.1	1.7
15.0	2	0	0.0	0.0	0.0	0.0

Milepost	Severity	Estimated Length by Raters	Actual Length, All	2-Point Length	Difference	Percent Difference
15.1	2	0	0.0	0.0	0.0	0.0
15.2	2	0	0.0	0.0	0.0	0.0
129.0	2	0	0.0	0.0	0.0	0.0
129.1	2	8	0.0	0.0	0.0	0.0
129.2	2	0	0.0	0.0	0.0	0.0
129.3	2	0	0.0	0.0	0.0	0.0
144.8	2	40	55.7	51.5	4.2	7.5
144.9	2	84	126.3	109.2	17.1	13.5
19.0	2	54	133.5	122.7	10.8	8.1
19.1	2	24	39.3	36.9	2.4	6.1
19.2	2	0	0.0	0.0	0.0	0.0
17.0	3	0	0.0	0.0	0.0	0.0
17.1	3	0	0.0	0.0	0.0	0.0
17.2	3	0	0.0	0.0	0.0	0.0
15.0	3	0	0.0	0.0	0.0	0.0
15.1	3	0	0.0	0.0	0.0	0.0
15.2	3	0	0.0	0.0	0.0	0.0
129.0	3	0	0.0	0.0	0.0	0.0
129.1	3	0	0.0	0.0	0.0	0.0
129.2	3	0	0.0	0.0	0.0	0.0
129.3	3	0	0.0	0.0	0.0	0.0
144.8	3	16	5.7	5.3	0.4	7.0
144.9	3	28	9.4	9.4	0.0	0.0
19.0	3	0	0.0	0.0	0.0	0.0
19.1	3	0	0.0	0.0	0.0	0.0
19.2	3	0	0.0	0.0	0.0	0.0

Table 11: Transverse Cracking 2-Point and Actual Length Field Measurements, Continued

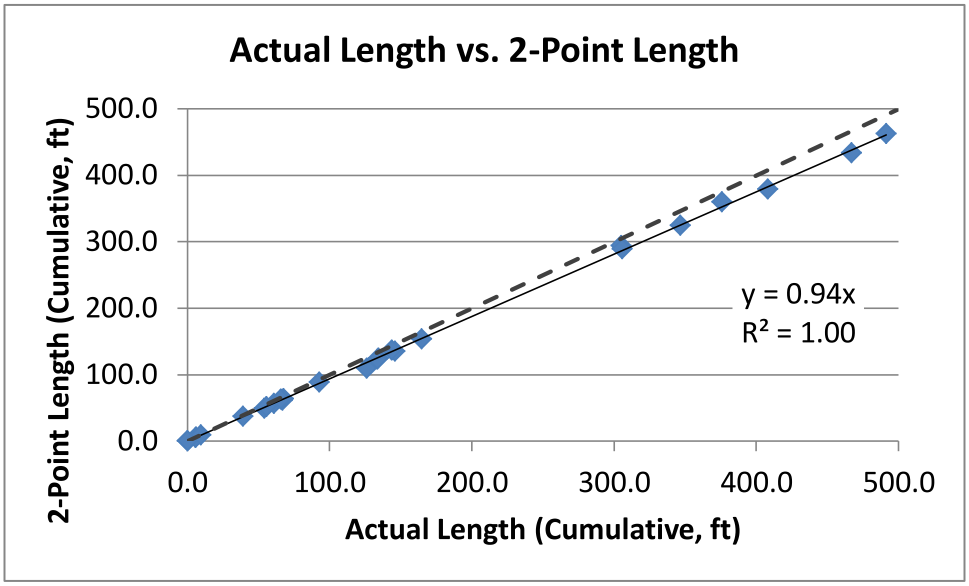


Figure 43: Actual Length vs. 2-Point Length, Cumulative Transverse Cracking

The mean and standard deviation of the length difference were 6 ft and 8.8 ft, respectively. The mean and standard deviation of the percent difference were 3% and 3.7%, respectively. Considering that the HPMS and PMS parameters are estimated from raters' data and that the purposes of the visual surveys and both HPMS and PMS data are to evaluate pavement condition at the network level, the difference found between actual and 2-point length is not significant in practice and should not be a concern in regard to the quality of the information estimated from visual distress surveys.

4.3.3 Alligator Cracking Analysis

The new protocol for visual distress surveys requires that pavement evaluators report their pace count of every instance of alligator cracking found within each test section and also whether the cracking is present in one or two wheelpaths. This rating includes longitudinal cracking present within the wheelpaths, as is a common industry practice.

For each severity of cracking, as shown previously, the area of the test section affected is calculated by multiplying the number of paces by the evaluator's pace length by the number of wheelpaths by the assumed wheelpath width of 2 feet. Field measurements were taken from the same mileposts used in the transverse cracking analysis. Table 12 shows the results from the detailed field measurements.

Table 12: Field Measurements for Alligator Cracking Analysis

Route	Milepost	Severity Rating	Area by Raters	Measured Area	Comments
NM 28	17.0 P	1	25.5	75.3	
NM 28	17.1 P	1	195.5	160.2	
NM 28	17.2 P	1	81.2	211.8	
I 25	15.0 P	1	534.1	347.5	Discrepancy - omitted.
I 25	15.1 P	1	951.3	1033.3	
I 25	15.2 P	1	349.5	461.8	
I 10	129.0 P	1	50.3	33.5	
I 10	129.1 P	1	0.0	0.0	
I 10	129.2 P	1	0.0	0.0	
I 10	129.3 P	1	0.0	0.0	
US 70	144.8 M	1	488.8	739.5	Area much wider than 2 feet.
US 70	144.9 M	1	255.0	981.7	Area much wider than 2 feet.
NM 478	19.0 P	1	192.3	239.5	
NM 478	19.1 P	1	406.5	562.7	
NM 478	19.2 P	1	274.4	416.5	
NM 28	17.0 P	2	1.8	14.7	
NM 28	17.1 P	2	64.3	162.5	
NM 28	17.2 P	2	364.7	341.1	
I 25	15.0 P	2	0.0	168.2	Discrepancy - omitted.
I 25	15.1 P	2	0.0	61.5	
I 25	15.2 P	2	14.4	19.9	
I 10	129.0 P	2	57.3	45.9	
I 10	129.1 P	2	0.0	0.0	
I 10	129.2 P	2	0.0	0.0	
I 10	129.3 P	2	0.0	0.0	

Route	Milepost	Severity Rating	Area by Raters	Measured Area	Comments
US 70	144.8 M	2	62.4	196.1	Area much wider than 2 feet.
US 70	144.9 M	2	61.0	219.8	Area much wider than 2 feet.
NM 478	19.0 P	2	0.0	23.5	
NM 478	19.1 P	2	0.0	2.3	
NM 478	19.2 P	2	0.0	0.0	

Table 12: Field Measurements for Alligator Cracking Analysis, Continued

Two data points were omitted from the analysis because they showed discrepancies in the values obtained by raters and by measurements. The three raters reported longitudinal cracks along the wheel path of severity 1 only. In a second look and by consensus, some of these cracks were upgraded to severity 2 when measured and recorded in the field tests. The sum of the alligator cracking area of severities 1 and 2 from field tests was 515.7 ft^2, which is very close to the average area of alligator cracking from the three raters (534.1 ft^2).

It is important to note that these test sections did not contain any alligator cracking of severity 3. In some cases, instances of alligator cracking of severity 2 or higher can have a width greater than 2 feet, which is the assumed width recommended by FHWA in calculating the area affected. This was particularly evident in the measurements taken on US 70 and also indicated in the results of the analysis, which shows that the raters' evaluations underestimate the area of alligator cracking present in the test section by approximately 30%, as demonstrated in Figure 44. The dashed line represents the perfect fit (y = x).

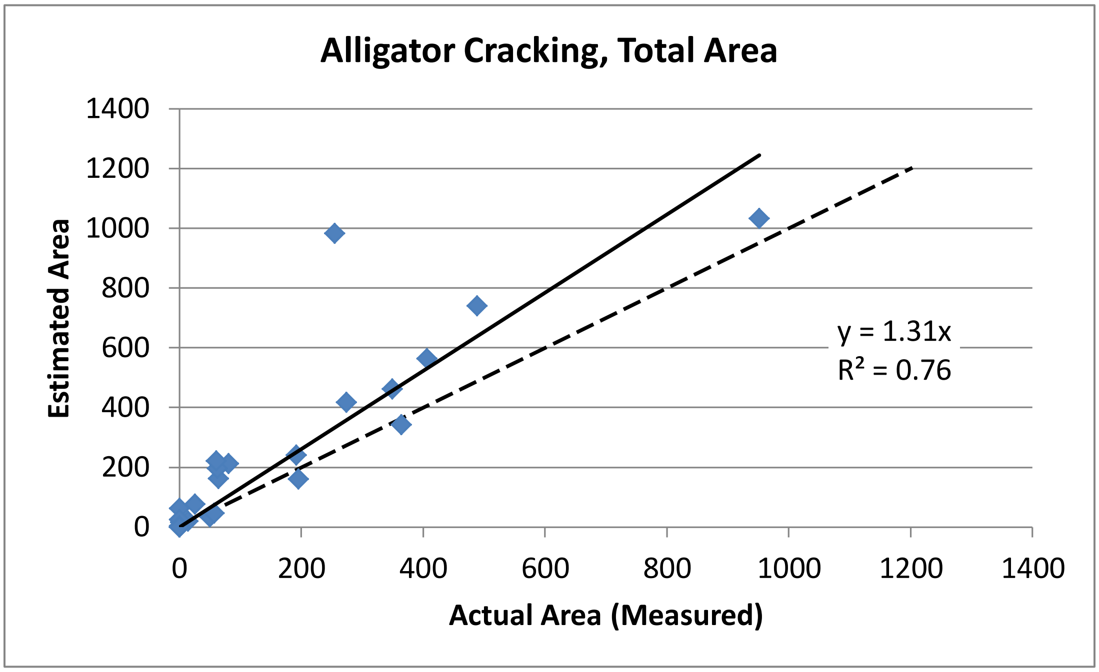


Figure 44: Alligator Cracking, Estimated Area vs. Detailed Measurements

Given the results of this analysis, the HPMS reporting data can be obtained by multiplying the calculated area of alligator cracking by a factor of 1.30 to account for the underestimation of the true area by raters' pacing methods.

4.4 Distress Rate Analysis

The analysis of the distress rate requires a few base assumptions because the distresses that are collected manually have changed between the old and new protocols. Previously, the distress rate was based on eight distresses that were all collected manually from sample sections. Now, two distresses have been eliminated from that group – rutting and shoving will be collected using a laser profilometer and, and patching has been eliminated altogether, the reasoning being that patching is not a maintenance solution for the types of roads that are rated for HPMS purposes. Another change is that longitudinal cracking has been combined with edge cracking to form one category, leaving only 5 distresses that are collected manually. The assumptions listed below are those required to form a direct comparison between the old and new methods:

- Rutting and shoving and patching will be assumed to remain constant. The same values will be used in both analyses.
- Because longitudinal and edge cracking have been combined, the same values will be used for these distresses, although their respective weights will be used.
- Only the extent of the worst severity reported will be used, since the old protocol only calls for worst severity present to be rated.

The comparison of the distress rates in the old and new methods between rounds (i.e. Old Method Round 1 vs. Old Method Round 2) are not affected by these assumptions.

The data collected from all mileposts was entered into the distress rate formula and compared among rounds. Figure 45 shows the correlation between rounds for the old protocol. The relationship between rounds is very nearly equal (y = x). This suggests that the data collected in the old protocol does not contribute to extreme variation in the distress rate between ratings

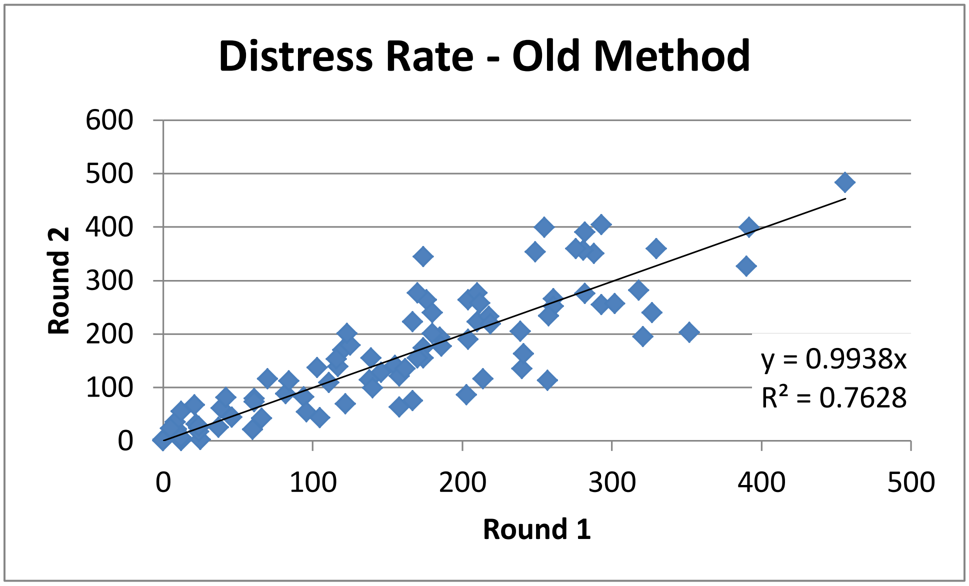


Figure 45: Distress Rate Comparison – Old Method

The new protocol collects data by counting the number of transverse cracks and calculates percent area for alligator cracking, therefore, the data must be translated into the same terms as the old protocol (severity and extents) to be used in the distress rate formula. As shown previously, methods have been developed that convert these two distresses into severity and extent. Figure 46 shows the relationship between rounds of the new protocol's distress rates. The relation is nearly one to one in this case also, suggesting that the distress rate is not sensitive to methods of data collection within the new protocol as well and can produce consistent distress rates to be used in calculating the PSI, which is important on several levels for management and federal reporting.

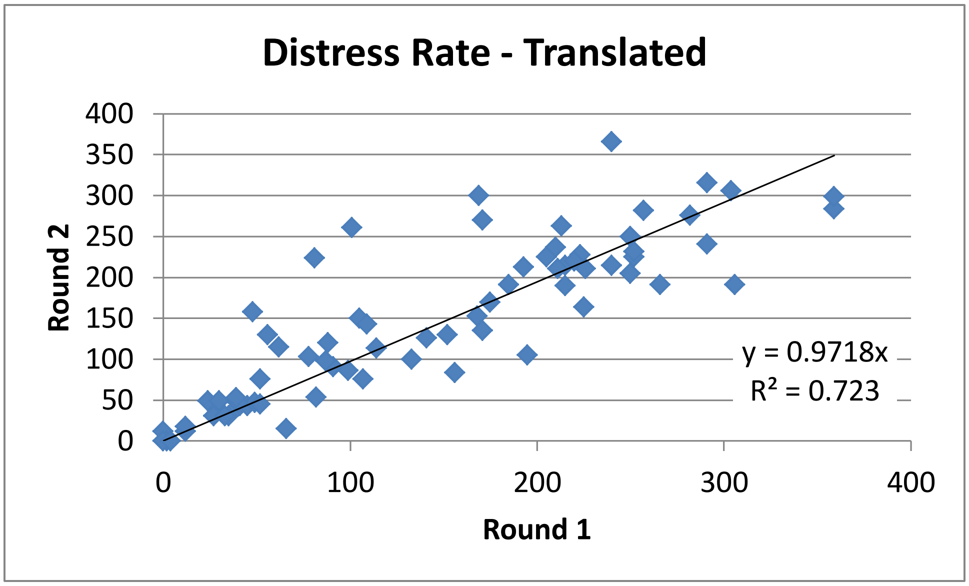


Figure 46: Distress Rate Comparison: Translated New Protocol

While distress rates are consistent among old and new protocols, it is also imperative to compare the distress rate between the methods to be able to evaluate the consistency of important measures after the switch to the new protocol. Since the distress rate is used in New Mexico's PSI, if there are large changes, the reported conditions of the network could drastically change within one reporting cycle, creating a potential problem regarding legislation and budgetary needs. Figures 47 and 48 show the results of the new protocol compared to the old protocol in round 1 and round 2, respectively. The results are also statistically acceptable, indicating that the translation between methods and adopting the new method of data collection will not adversely affect reporting requirements or change the PSI values dramatically.

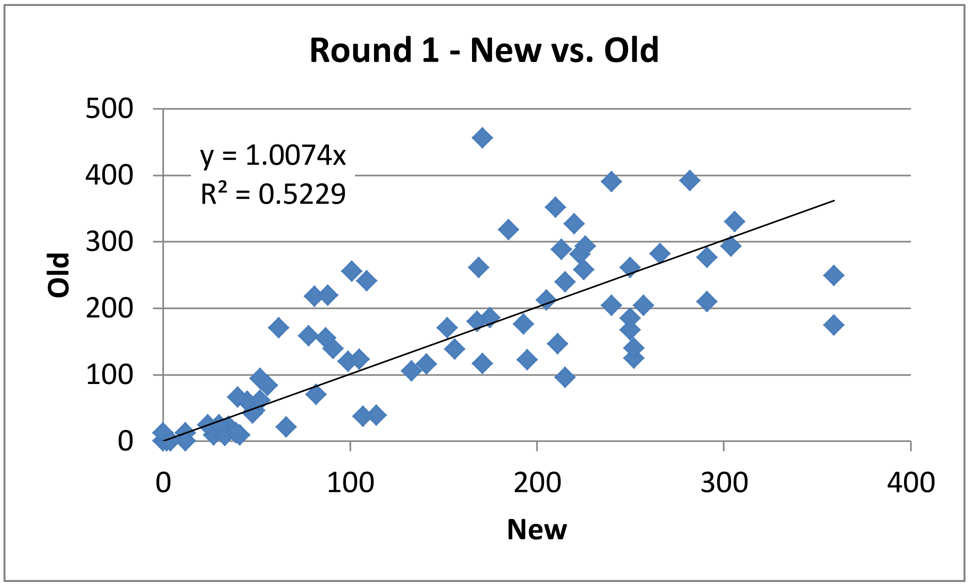


Figure 47: Distress Rate Comparison – New vs. Old Protocols – Round 1

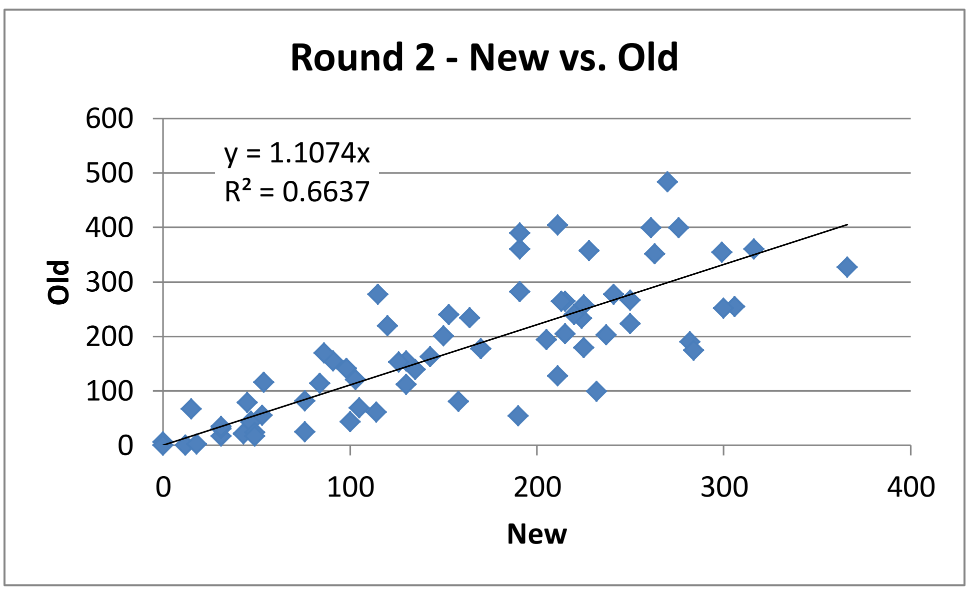


Figure 48: Distress Rate Comparison – New vs. Old Protocols – Round 2

4.5 Time Required for Visual Surveys

The new protocol for flexible pavements evaluates five distress types. For raveling and weathering, raters report the highest severity rating only. The extent of the distress of raveling and weathering is assumed to be 3 (high), indicating the whole section is affected. For bleeding, the new protocol rates all severity levels present and assumes that the extent of each severity found is 1 (Low). For these two distresses, the time required to rate and report the information is about the same as in the old protocol.

For the transverse, longitudinal and alligator cracking, the new protocol requires that the raters report all severity levels and the corresponding extent of those severities, given in either paces (or pace counts) or extent. The raters will have to keep track of pace and percentage of section for several distress types and severity levels. For sample sections having all or most distress types considered by the new protocol and in fair or poor condition, the time required for the survey may be longer compared to the old protocol. The required survey time per section also may vary with the level of experience of the raters. However, in the trial runs performed by UNM and NMSU crews, the time required to evaluate a given section was comparable for both protocols.

Data collected by Rater 1 of UNM were analyzed for time to complete the field rating. Rater 1 was the only evaluator who had listed times on every milepost and, therefore, had the most complete data set regarding time. An additional column in Table 13 was added to compare the distress rate with the time required to rate the section. Overall, no correlation was found between the distress rate and the time it takes a rater to evaluate a section. The expectation was to find higher distress rates in mileposts that took the longest to rate, but this did not happen. This could have occurred because of the way the distress rate is calculated, which places much heavier weights on certain distresses while virtually ignoring others. The average time that it took Rater 1 to evaluate test sections in the old and new protocols was very close, suggesting that the time it takes to apply these protocols do not differ as much as previously thought. In most cases, the new protocol actually took less time to complete, but it is important to note that by the time of implementation of the new protocol, the rater had already had some practice rating sections at least twice, so the experience could have played a role. The more experience the raters have, the faster they are able to rate sections. The summary of the rating times is shown in Table 13.

Table 13: Rater 1 – Evaluation Times

		Distress Rate - Old	**Distress Rate - New**	**Time to Rate (min)**	
Route	**MP**			**Old**	**New**
NM0041	0-P	43	41	12	7
	1-P	47	49	14	7
	2-P	91	91	15	11
	3-P	86	99	10	12
	29-P	215	240	10	11
	30-P	282	257	8	7
	31-P	284	359	11	11
	32-P	241	291	10	9
NM0014	0-P	205	250	10	8
	1-P	250	250	10	8
	2-P	250	250	8	8
	3-P	215	215	11	11
US0550	0-P	45	52	11	9
	1-P	105	195	8	7
	2-P	191	185	11	10
	3-P	211	211	11	10

		Distress Rate - Old	Distress Rate - New	Time to Rate (min)	
Route	**MP**			**Old**	**New**
NM0556	12-M	31	27	9	9
	13-M	12	12	9	9
	14-M	49	24	13	10
	15-M	0	2	10	6
NM0006	0-P	306	304	11	9
	1-P	115	62	13	11
	2-P	130	152	11	9
	3-P	153	158	13	11
Average Time				10.8	9.2

Table 13: Rater 1 – Evaluation Times, Continued

Chapter 5

Conclusions

5.1 Conclusions

This study compared two methods of manual data collection of pavement distress. The "old" protocol was the method currently in place in New Mexico's Department of Transportation Pavement Management System prior to this study, up to 2011. This method used a severity and extent structure that reports the highest severity along with the extent of that severity in a 0, 1, 2,or 3 format. The format that distress data had been previously reported according to the "old" protocol was not useful for the needs of the Federal Highway Administration's Highway Performance Monitoring System, a federally mandated pavement management system. The latter requires annual reporting of specific distress measures. As a result, a new protocol was developed and tested for use in the HPMS and the state of New Mexico's Pavement Management System. The new protocol was tested for data reliability among several raters using interrater reliability processes. This protocol was also tested using the same processes of a given rater over time to determine repeatability of this new method through average deviation analysis. Overall, the data collected using the new format are more consistent than the old protocol and serve the purpose of meeting both agencies' needs regarding data reporting and minimal disruption of legacy data flow.

5.1.1 Observations

Although the pavement evaluators who participated in the data collection were relatively inexperienced, the ranges of acceptability were met in almost every distress analyzed. Overall, the general trend for the data collection analyses shows a decreasing variability across rounds, regardless of the method used. This is important to note because it demonstrates the value of training and experience with regards to data consistency.

It is worth mentioning also that the state of New Mexico adopted the new protocol as of summer 2012. The data will be collected from May through August.

5.1.2 Time-History Data Collection Issues

One of the major challenges of successfully implementing and maintaining a PMS is ensuring consistency of legacy data when new techniques and technologies are implemented. More often than not, pavement management data are stored in several databases that use different identifiers and terms within a single agency. These legacy systems can have unique identifiers that are difficult or impossible to update to current technology and systems (AASHTO, 2001). Compatibility of the pavement condition data collection over time is very important for supporting effective pavement management. Quality time-series of pavement condition data are needed to develop reliable deterioration models, measure the impact of maintenance and rehabilitation treatments, develop multi-year work plans, and optimize the allocation of resources (Flintsch & McGhee, 2009). Therefore, it is important that the new and legacy data are compatible or can be made compatible through an appropriate conversion. This applies to the actual data attributes (e.g., type of crack and length) and to the location referencing. The use of appropriate metadata can facilitate the transition. The issue of ensuring consistency over time is particularly important at the onset of adopting automated technologies. This typically creates significant challenges in terms of ensuring that the criteria and metadata are properly referenced.

5.1.3 Pavement Condition Data Consistency

The first concern with the adoption of a new data collection technology or with the contracting of a service provider is the verification that the pavement measurements are at least as accurate as the existing data and consistent with agency protocols and requirements (NCHRP, 2009). Furthermore, it is also important that the new data can be processed to provide pavement condition indicators that are consistent with the agency's historical data to allow time-history analyses. For example, it is important that automated crack detection systems provide the same results as the agency's visual method. Verification tests could be included in the quality management programs to verify this agreement. Several DOTs have used a pre-qualification process, in which the agencies ask potential service providers to conduct measurements on several control sections for which the agency has conducted reference measurements. Another example is the certification process that has been proposed for profilers. Verification of the consistency of the data is also important when changing service providers or when the service providers (or the agency itself) use more than one pavement data collection piece of equipment or technology. The main reason that most agencies resist changes in data collection methods is that their banks of legacy data may no longer be useful (McGhee, 2004).

5.2 Implementation

The new protocol was implemented in the summer of 2012 by the NMDOT. The author of this thesis provided a week-long training event for the students at both UNM and NMSU, with the assistance of the Principal Investigators of the project, Dr. Susan Bogus Halter (UNM) and Dr. Paola Bandini (NMSU), and Mr. Robert Young from the NMDOT. This training covered the pavement distresses, causes, rating criteria, and practice sessions in the field. The week following the training, the universities took over and did more practice sessions and provided training on office and clerical procedures, scheduling, and procedures that are required by each university. The training is done typically one or two weeks after finals week, so that the collection process can immediately begin to ensure the required routes are collected within the appropriate time frame over the summer.

The universities were required to collect data on all of the NHS routes and one-half of the non-NHS routes in the New Mexico State Highway System (about 10,222 data collection sites) assigned to the universities based on location, northern and southern parts of the state. However, because of better than anticipated productivity on the part of the data collection crews, the universities collected data on 15,292 data collection sites, nearly the entire New Mexico State Highway System.

UNM and NMSU completed the evaluation of the New Mexico State Highway System in August 2012, fulfilling the terms and conditions of their contracts. As a result, the Department is in compliance with State and Federal laws. In each of the five years when UNM and NMSU were contracted for distress data collection since 2006, the universities submitted the required data on time, under budget, and with a perfect safety record. There were no reportable problems or negative incidents during the distress data collection activities. This program has been very successful in terms of cost, data quality and deliverables, and timely completion.

The NMDOT received all necessary data from the universities for State highway system oversight and management up to 2012. The data is being loaded into the Department's upgraded Pavement Management System (PMS) which will provide the Department's decision makers with a factual basis for making investment decisions. These decisions will affect many millions of dollars of taxpayer money and the information needed to determine the needs of the network as a whole. The data will also aid in formulating statewide programs of new construction, rehabilitation and maintenance programs to optimize the use of scarce state and federal funds over the New Mexico State Highway System.

The results of the Quality Assurance (QA) program are being finalized but the data is of good quality. The Department's quality checks require 5% of the total student data points be evaluated by Department personnel that was trained at the same time as the students. Of 483 sites where Quality Assurance checks were conducted, there were only 32 sites across both schools where one of the seven distresses measured by the corresponding QA Inspector was inconsistent with the data collected by the universities as per the minimum acceptable limits of variability defined in the Department's Pavement Distress Data Quality Management Plan (less than 1%). At the time of this thesis, it is unclear whether the pavement evaluators or the QA personnel were having the most difficulty rating sections. These 32 sites are being re-examined by Independent Verification Inspectors to determine the reasons for the inconsistencies. Upon first look, it seems that raters are having the most difficulty with rating bleeding, which was expected due to the nature of the distress or rating criteria.

Each university was asked to provide feedback on the new data collection methods. These suggestions were provided by Robert Young (personal email, March 2012).

UNM has provided the following recommendations for improving the Pavement Distress Data Collection Program:

1. The student evaluators suggested that more time be spent on field training and less with slide presentations, especially rating concerns with cracking in locations with curb and gutter and bleeding. Additional field training should improve agreement on evaluations among the crew members. The training time spent on performing an evaluation of rigid pavement could be reduced. In addition, it was suggested that more time be spent on explaining the information contained in the reference books, including the route termini and route lists. This should improve the amount of miles evaluated per day, as evaluators will spend less time trying to find a specific route.
2. The NMDOT's Access database pavement condition software provided evaluators with a graphical interface to enter distress data. Each evaluation crew entered into the electronic forms stored in their portable computers their daily distress evaluations, which were then retrieved in a text file on a weekly basis. The text files contained all of the evaluations to date for a given crew. The 2012 UNM evaluation team made suggestions on the issue of data entry. These included changing the way exceptions forms work, in order to allow the evaluators to edit

them, when necessary. Because of this problem, some exceptions forms were duplicated in the text files. Another issue that was found by the UNM evaluation team was the fact that NMDOT did not update the database with the milepost exceptions and changes recommended in 2009, causing confusion in some areas, due to contradictions among the database, the route termini, and master route list. Lastly, the students addressed the need for a more keyboard-friendly interface. They feel that the latter will let them enter the data more easily, especially when they are in the field.

NMSU has provided the following recommendations for improving the Pavement Distress Data Collection Program:

1. The NMDOT's Access database pavement condition software was used by the raters to input the distress data. The software produces a text file with the distress up-to-date data uploaded in a given portable computer. The part of the software to create or edit exceptions (i.e., when test sections are not in the database) was found to have problems. Once an exception was created, it was not possible to edit the data for this section or delete the exception from the computer. Also, the location and pavement type information shown in the exception files did not correspond to the respective test sections. To avoid confusions, the NMSU team had to correct these errors manually before the files could be submitted to the NMDOT Project Manager.
2. The master database in the NMDOT software does not contain a number of test sections that were assigned to the crews. These test sections had to be input in the software as "exceptions." It is recommended that the test sections reported in the

exception files that are permanent be incorporated in the master database. Also, duplicated test sections (i.e., those that appear in both flexible and rigid databases) should be corrected in the master database. Exceptions due to permanent features in the highway (e.g. ramp, weight station or bridge within a given test section) should be corrected in the master database. Also, test sections in unpaved roads should be deleted from the database. Information provided by the raters in the "Remarks" field of the data forms can also be used to update the database. These corrections will significantly reduce the work of the crews regarding uploading data and reduce or eliminate errors and confusion in the process.

3. Regarding the rating criteria for flexible pavements, two raters suggested that Bleeding should be rated with both severity and extent. They found test sections with a given severity level in more than 30% of the section area. Currently, all the severity levels of Bleeding are reported, and the extent of each severity level is not rated or reported. This part of the criterion may deserve further discussion.
4. It is recommended to clarify the rating criteria for Longitudinal Cracking and Edge Cracking, particularly for test sections with curb and gutter.
5. It will be useful for the NMSU team to receive from NMDOT the road maintenance and construction schedule for the data collection period within the NMSU area before the field work starts.
6. In addition to the data input by the raters, the text files for flexible test sections contained data calculated or deduced by the software. It was found that the software writes "3" in the columns labeled as severity of Alligator Cracking and

Transverse Cracking when the severity and extent of these distresses are zero. This part of the software needs to be reviewed carefully.

Because this was the first year that the new protocol was implemented, we expected that there would be bugs that needed correction. The distress data software system had to be entirely reprogrammed to accommodate the new data formats. The testing time available to test the software was short. The database issues may take more time to fix, as the NMDOT has several databases that interact with each other in the PMS. In all, the program carried out by the universities was a success, the data variability was reduced, and the productivity was higher than in the previous data collection cycles despite the small problems encountered.

5.3 Future Research

It would be beneficial to implement the new protocol for a few more years and among several states to determine whether the detailed measurement correction factors apply to all pavement sections, or if the few test sections that were measured in detail were outliers in the research. Any method can be refined, and it may take more than one project to define these factors.

Advances in technology occur at almost an exponential rate, and the improvements in fully automated data collection are no exception. 3-D modeling vans that use Lidar technology have been in use since 1997 and had trouble in detecting cracks filled with sand, and had trouble detecting roughness due to raveling and bleeding of pavements (Bursanescu and Blais, 1997). However, even the state-of-the-art distress logging vans in use today still have trouble with some basic surface distresses such as raveling and bleeding, while their crack detection systems are quite accurate. It is recommended as a follow up to this research to determine whether an automated data distress collection system can produce the same results of same or better quality as a manual process, including severity and extent of all distresses present on a test section.

Appendices

Appendix A – Old Field Form

UNM/NMDOT Flexible Pavement Evaluation

Flexible Pavement Condition Evaluation Form, 2006

Posted Route: ____________ Milepost: ____________ Direction: M P

Pavement Distress Type:		Severity				Extent			
Surface Defects	Ravel & Weathering:	1	2	3	N	1	2	3	N
	Bleeding:	1	2	3	N	1	2	3	N
Surface Deformations	Rutting & Shoving:	1	2	3	N	1	2	3	N
Cracking Distresses	Longitudinal Cracking:	1	2	3	N	1	2	3	N
	Transverse Full Width Cracking:	1	2	3	N	1	2	3	N
	Alligator Cracking:	1	2	3	N	1	2	3	N
	Edge Cracks:	1	2	3	N	1	2	3	N
Maintenance Patching	Patch Condition:	1	2	3	N	1	2	3	N

Remarks: ____________________

Date: ____________ Inspector: ____________

Posted Route: ____________ Milepost: ____________ Direction: M P

Pavement Distress Type:		Severity				Extent			
Surface Defects	Ravel & Weathering:	1	2	3	N	1	2	3	N
	Bleeding:	1	2	3	N	1	2	3	N
Surface Deformations	Rutting & Shoving:	1	2	3	N	1	2	3	N
Cracking Distresses	Longitudinal Cracking:	1	2	3	N	1	2	3	N
	Transverse Full Width Cracking:	1	2	3	N	1	2	3	N
	Alligator Cracking:	1	2	3	N	1	2	3	N
	Edge Cracks:	1	2	3	N	1	2	3	N
Maintenance Patching	Patch Condition:	1	2	3	N	1	2	3	N

Remarks: ____________________

Date: ____________ Inspector: ____________

Posted Route: ____________ Milepost: ____________ Direction: M P

Pavement Distress Type:		Severity				Extent			
Surface Defects	Ravel & Weathering:	1	2	3	N	1	2	3	N
	Bleeding:	1	2	3	N	1	2	3	N
Surface Deformations	Rutting & Shoving:	1	2	3	N	1	2	3	N
Cracking Distresses	Longitudinal Cracking:	1	2	3	N	1	2	3	N
	Transverse Full Width Cracking:	1	2	3	N	1	2	3	N
	Alligator Cracking:	1	2	3	N	1	2	3	N
	Edge Cracks:	1	2	3	N	1	2	3	N
Maintenance Patching	Patch Condition:	1	2	3	N	1	2	3	N

Remarks: ____________________

Date: ____________ Inspector: ____________

Appendix B – Old Rating Criteria

Flexible Pavement: Pavement Evaluation Reference Chart

DISTRESS	SEVERITY	EXTENT
Raveling & Weathering: The wearing away of the pavement surface, due to dislodged aggregate particles and loss of asphalt binder. Normally the distress will be throughout the test section.	Low (1): Aggregate or binder has started to wear away on pavement surface. Some dislodged aggregate can be found on the shoulder. Med (2): Aggregate or binder has worn away. Surface texture is rough and pitted. High (3): Aggregate and/or binder has worn away, and surface appears is severely rough and pitted.	Low (1): 1% to 20% of test section. Med (2): 21% to 60% of test section. High (3): 61% of test section, or more.
Bleeding: A film of bituminous material on the pavement surface.	Low (1): Film is evident, but aggregate can still be seen. Spotty. Med (2): Film is clearly seen, covers most of the aggregate, and is a little sticky. High (3): Film is predominant, very sticky, and material is thick enough to show.	Low (1): 1% to 20% of test section. Med (2): 21% to 60% of test section. High (3): 61% of test section, or more.
Rutting and Shoving: Longitudinal surface depressions in wheel path. (Check with a 4-foot rut bar.)	Low (1): ¼-inch to ½-inch in depth. Med (2): ½-inch to 1-inch in depth. High (3): More than 1-inch in depth.	Low (1): 1% to 20% of test section. Med (2): 21% to 60% of test section. High (3): 61% of test section, or more.
Cracks: **Longitudinal Cracks:** Wheel Track Mid-Lane Center Line **Transverse Cracks:** Full Width	Low (1): Sealed or non-sealed with a mean width of less than ¼-inch. May have very minor spalls. Med (2): A. Sealed or non-sealed, and moderately spalled. Any width. B. Sealed, but sealant separated allowing water to penetrate. C. Non-sealed cracks that are not spalled, but are over ¼-inch wide. D. Low severity alligator cracks near crack, or at the corners of intersecting cracks. E. Causes a significant bump to a vehicle. High (3): A. Severely spalled. (Any width.) B. Medium to high severity alligator cracks exist near the crack, or at the corners of intersecting cracks. C. Causes a severe bump to a vehicle.	Low (1): 1% to 20% of test section. Med (2): 21% to 60% of test section. High (3): 61% of test section, or more.
Alligator Cracks: Pattern of interconnected cracks resembling chicken wire or alligator skin.	Low (1): Hairline, disconnected cracks 1/8-inch wide, or less. No spalls. Med (2): Fully developed crack pattern greater than 1/8-inch wide. Lightly spalled. High (3): Severely spalled. Cells rock. May pump.	Low (1): 1% to 20% of test section. Med (2): 21% to 60% of test section. High (3): 61% of test section, or more.
Edge Cracks: Cracks which occur on the edge of the pavement.	Low (1): ¼-inch wide, or less. No spalls. Med (2): Greater than ¼-inch wide. Some spalls. High (3): Severely spalled.	Low (1): 1% to 20% of test section. Med (2): 21% to 60% of test section. High (3): 61% of test section, or more.
Patching: An area where the original pavement has been removed and replaced with similar or different material. Types of Patching: Hot Mix Patch. Skin Patch. Other types (Please note on "notes section" of the evaluation card.)	Low (1): Patch is present, and is in good condition. Med (2): Somewhat deteriorated. Low to medium of any type of distress on patch. High (3): Patch is deteriorated to point of soon or immediately needing replacement.	Low (1): 1% to 20% of test section. Med (2): 21% to 60% of test section. High (3): 61% of test section, or more.

Appendix C – New Field Form

New Mexico Department of Transportation

Distress Rating of Flexible Pavements

Pace (ft): ______

Route: ______

Milepost:			Direction:	M	P
Raveling & Weathering Circle Highest Severity Present		0	1	2	3
Bleeding Circle All Severities Present		0	1	2	3
Alligator Cracking Single crack in Wheel Path is Severity 1, mark Severity 0 if section not cracked Length is in # Paces, indicate 1 or 2 wheelpaths	**Severity**	**Length (1 or 2) [e.g., 43 (2)]**			
	0				
	1				
	2				
	3				
Transverse Cracking (Any crack >6' is counted) Count all cracks present, mark Severity 0 if section not cracked.	**Severity**	**Counts**			**Total**
	0				
	1				
	2				
	3				
Longitudinal Cracking (ANY outside the wheelpath) Estimate % area affected by EACH severity	**Severity**	**Extent**			
	0	0	1	2	3
	1	0	1	2	3
	2	0	1	2	3
	3	0	1	2	3

Remarks: ______________________________

Date: ______________ Inspector: ______________

Appendix D – New Rating Criteria

Flexible Pavement: Pavement Evaluation Reference Chart

DISTRESS	SEVERITY	NOTES
Raveling & Weathering: The wearing away of the pavement surface, due to dislodged aggregate particles and loss of asphalt binder.	(1) Low: Aggregate or binder has started to wear away. (2) Medium: Aggregate or binder has worn away. Surface texture is rough. Some dislodged aggregate can be found on the shoulder. (3) High: Aggregate and/or binder has worn away, and surface texture is severely rough and pitted.	MOST PREVALENT severity
Bleeding: A film of bituminous material on the pavement surface, from the asphalt concrete mix and not from the vehicles or external sources.	(1) Low: Film is evident, but aggregate can still be seen. Spotty. (2) Medium: Film is clearly seen, covers most of the aggregate and is a little sticky. (3) High: Film is predominant, very sticky, and material is thick enough to shove.	Mark EACH severity present.
Transverse Cracks: Half-width lane to full-width lane transverse cracks (6ft or longer cracks). Disregard cracks shorter than 6 ft.	(1) Low: Unsealed, mean width of less than ¼-inch. OR sealed with sealant in good condition, any width. (2) Medium: Any crack with mean width greater ¼-inch and less than ¾ inch. May have adjacent Low severity random cracks; some spalling. (3) High: Any crack wider than ¾ inch, may have adjacent moderate to high random cracking.	COUNT the number of cracks of EACH severity within sample section. Include half-width lane and longer cracks. Record the totals of each severity on field form.
Alligator Cracks: Pattern of interconnected cracks resembling chicken wire or alligator skin. Longitudinal cracks in the wheel path are rated as Low severity alligator cracking. Severities 2 and 3 must have at least 3 cells.	(1) Low: Hairline, disconnected cracks, 1/8-inch wide or less, less than 3 cells. No spalls. AND/OR a longitudinal crack, any severity, in the wheel path. (2) Medium: Fully developed cracks greater than 1/8-inch wide. Three or more cells. Lightly spalled. (3) High: Severely spalled, cells rock, and may pump.	PACE OFF the cumulative lengths of EACH severity present. Record lengths (in paces). Mark location of occurrence in field form: 1 or 2 wheel paths
Edge Cracks: Cracks that lie within 1 foot of the edge line. Does NOT apply in roads with curb and gutter installations. DO NOT rate the shoulder.	(1) Low: Less than ¼-inch wide. No spalls. (2) Med: Greater than ¼-inch wide. Some spalling may be present, but pavement is still intact. (3) High: Severely spalled. Pieces of pavement have broken off the edge of the roadway.	(1) Low: 1% to 30% of test section. (2) Med: 31% to 60% of test section. (3) High: 61% of test section, or more.
Longitudinal Cracks: ANY longitudinal crack NOT in the wheel path, but NOT within 1' of the edge line.	(1) Low: Unsealed, mean width of less than ¼-inch. OR sealed with sealant in good condition, any width. (2) Medium: Any crack with average width greater than ¼-inch and less than ¾ inch. May have adjacent Low severity random cracks and some spalling. (3) High: Any crack wider than ¾ inch, may have adjacent moderate to high random cracking and spalling.	(1) Low: 1% to 30% of sample section. (2) Medium: 31% to 60% of sample section. (3) High: 61% or more of sample section.
Patching: Any new pavement placed into the pavement section. Extent is rated as percent of the test section affected.	(1) Low: Patch is in good condition. (2) Medium: Somewhat deteriorated, has Low to Medium severities of any distress present. (3) High: Needs replacement. High Severity of any distress, gaps are present between the pavement and the patch.	(1) Low: 1% to 30% (2) Med: 31% to 60% (3) High: 61% or more

10% Rule: If 10% or more of the distress shows a higher severity, use this higher severity to rate the distress.

References

American Association of State Highway Officials (1961) *AASHO Road Test: History and Description of Project. Publication 816, National Research Council, National Academy of Sciences*, Washington, D.C.

American Association of State Highway and Transportation Officials. (1990). *Guidelines for Pavement Management Systems*. Washington, DC.

American Association of State Highway and Transportation Officials. (2001). *Pavement Management Guide* (p. 254). American Association of State Highway and Transportation Officials.

Amirkhanian, Serji N., Juang, C. Hsein, Koduru, Hari Krishan, and Xiao, Feipeng. *Using Fuzzy Logic and Expert System Approaches in Evaluating Flexible Pavement Distress: Case Study*. Journal of Transportation Engineering, Vol. 136, No. 2, February 2010, pp. 149-157.

Bandini, Paola, Bogus Halter, Susan, Montoya, Kelly, Pham, Hung V. and Migliaccio,Giovanni C. (2012). *Improving NMDOT's Pavement Distress Survey Methodology and Developing Correlations between FHWA's HPMS Distress Data and PMS Data.* New Mexico Department of Transportation Research Bureau in Cooperation with The U.S. Department of Transportation Federal Highway Administration, Albuquerque New Mexico.

Bandini, Paola, Bianchini, Alessandra, and Smith, David W. *Interrater Reliability of Manual Pavement Distress Evaluations.* Journal of Transportation Engineering, Vol. 136, No. 2, February 2010, pp. 165-172.

Bogus, Susan M., Lenke, Lary R., Song, Jongchul, and Waggerman, Raymond. *Rank Correlation Method for Evaluating Manual Pavement Distress Data Variability*. Journal of Infrastructure Systems, Vol. 16, No. 1, March 2010, pp. 66-72.

Bogus, S., Migliaccio, G., and Cordova, A. *Data Quality Assessment for Manual Pavement Distress Evaluations.* Proceedings of the Transportation Research Board Annual Meeting. Washington, DC, January 2010.

Bogus, S., Migliaccio, G., and Cordova, A. *Performance of Manual Condition Surveys Using Interrater Agreement Measurements.* Proceedings of the CIB World Congress 2010, Salford UK, May 2010.

Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). *On Average Deviation Indices for Estimating Interrater Agreement*. Organizational Research Methods, 2(1), 49.

Burke, M., & Dunlap, W. (2002). *Estimating Interrater Agreement with The Average Deviation Index: A User's Guide*. Organizational Research Methods, 5(2), 159. Res Methods Div.

Bursanescu, L., and Blais, F. *Automated Pavement Distress Data Collectionand Analysis: A 3-D Approach.* Proceedings of International Conference on RecentAdvances in 3-D Digital Imaging and Modeling. Ottawa, Ontario, Canada.May 12-15, 1997. pp. 311-317. NRC 41574.

Chua, Koon Meng and Xu, Ling. *Simple Procedure for Identifying Pavement Distresses from Video Images*. Journal of Transportation Engineering, Vol. 120, No. 3, May/June 1994, pp. 412-431.

Cordova-Alvidrez, Arturo. (2010) *A Framework for Assessing and Improving Quality of Data From Visual Evaluation of Asset Conditions*. Thesis, University of New Mexico.

Cronbach, L. (1990). *Essentials of Psychological Testing*. New York (5th.). New York: Harper-Row.

Daleiden, Jerome F. *Pavement Monitoring, Evaluation, and Data Storage* (2010). Transportation Research Board Online Publications, Accession Number 00784218. Accessed from http://onlinepubs.trb.org/epubs/millenium/00085.pdf, July 12, 2012.

Daleiden, Jerome F., and Simpson, Amy L. *"Off the Wall" Pavement Distress Variability Study* (1998).Transportation Research Record: Journal of the Transportation Research Board, 1643, 62-70.

Dunlap, William P.; Burke, Michael J.; Smith-Crowe, Kristin. *Accurate Tests of Statistical Significance for r[sub WG] and Average Deviation Interrater Agreement*. Journal of Applied Psychology, Apr 2003, Vol. 88 Issue 2, p356-262.

Ganesan, Venkatesa Prasanna Kumar. (2006) *Improving Pavements With Long-Term Pavement Performance: Products for Today and Tomorrow: Studyof Long-Term Pavement Performance (LTPP): Pavement Deflections ,Paper 2: Use of LTPP Data to Verify the Acceptance Limits Developed for Penndot Pavement Distress Data*. FHWA-HRT-06-109.

Gharaibeh, Nasir G., Zou, Yajie; and Saliminejad, Siamak. *Assessing the Agreement among Pavement Condition Indexes.* Journal of Transportation Engineering, Vol. 136, No. 8, August 2010, pp. 765-772.

Haas, Ralph. *Role of Standards in Pavement Management*. ASTM Standardization News, 1987, Vol 15 No. 4, pp 54-58.

Haas, Ralph, Hudson, W. Ronald, Zaniewski, John. *Modern Pavement Management* (1994). Krieger Publishing Co, Malabar, FL.

Haider, Syed Waqar, Chatti, Karim, Baladi, Gilbert Y., and Sivaneswaran, Nadarajah (2011). *Impact of Pavement Monitoring Frequency on Pavement Management System Decisions*. Transportation Research Record: Journal of the Transportation Research Board, 2225: 43-55.

Kaplan, R., & Saccuzzo, D. (1993). *Psychological Testing: Principles, Applications, and Issues.* Applications and Issues, 3rd ed. Brooks/Cole. Pacific Grove (3rd, p. 715). Cengage Learning.

Kulkarni, Ram B., Miller, Richard W. *Pavement Management Systems: Past, Present, and Future* (2003). Annual Meeting of the Transportation Research Board, National Research Council.

LeBreton, James M.; Senter, Jenell L. *Answers to 20 Questions about Interrater Reliability and Interrater Agreement*. Organizational Research Methods, Oct 2008, Vol. 11 Issue 4, p815-852.

Livneh M. (1994). *Repeatability and Reproducibility of Manual Pavement Distress Survey Methods*. Transportation Research Record: Third International Conference on Managing Pavements, 2: 279–289.

Miller, J. S., & Bellinger, W. Y. (2003). *Distress Identification Manual for the Long-Term Pavement Performance Program (FHWA-RD-03-031)*. Federal Highway Administration (4th revised.). Federal Highway Administration.

Miller, J., Rada, G., & Rogers, R. (1993). *Distress Identification Manual for the Long-Term Pavement Performance Project*. (SHRP) (3 ed., p. 147). Washington, DC: Strategic Highway Research Program, National Research Council.

Nevada Department of Transportation. *Flexible Pavement Distress Identification Manual*. State of Nevada Department of Transportation Materials Division. Carson City, NV,

New Mexico Department of Transportation. (2007). The New Mexico Department of Transportation's Pavement Evaluation Inspection Program: A Successful Partnership. Santa Fe, NM.

New Mexico Department of Transportation. (2007). The New Mexico Department of Transportation's Pavement Maintenance Manual. Santa Fe, NM.

New Mexico Department of Transportation. (2004). The New Mexico Department of Transportation’s Network Level Pavement Management. Santa Fe, NM.

Nunnally, J. (1978). *Psychometric Theory*. NY: McGraw-Hill Inc. New York: McGraw-Hill.

Oregon Department of Transportation. *Distress Survey Manual*. Salem, OR., 2007. www.oregon.gov/ODOT/HWY/CONSTRUCTION/docs/pavement/Distress_Survey_Manual.pdf.

Peshkin, D.G. and Hoerner, T. E. *Pavement Preservation: Practices, Research Plans, And Initiatives (*May 2005). National Cooperative Highway Research Program: Transportation Research Board. NCHRP Project No. 20-07, Task 184.

Peterson, Dale E. *Pavement Management Practices* (1987).. Transportation Research Board and National Cooperative Highway Research Program, Synthesis 135, Washington D.C.

Rada, G. R., Bhandari, R. K., Elkins, G. E., & Bellinger, W. Y. (1997). *Assessment of Long-Term Pavement Performance Program Manual Distress Data Variability: Bias and Precision.* Transportation Research Record: Journal of the Transportation Research Board, 1592, 151–168.

Rada, G. R., Wu, Chung L., Bhandari, R. K., Elkins, G. E., & Bellinger, W. Y. (1998). *Update of Long-Term Pavement Performance Manual Distress Data Variability – Bias and Precision.* Transportation Research Record: Journal of the Transportation Research Board, 1643, 71-79.

Shekharan, Raja, Frith, Douglas, Chowdhury, Tanveer, Larson, Charles, and Morian, Dennis (2007). *Effects of Comprehensive Quality Assurance/Quality Control Plan on Pavement Management*. Transportation Research Record: Journal of the Transportation Research Board, 1990: 65-71.

Smith, Roger E., Freeman, Thomas J., and Pendleton, Olga J (1998). *Contracting for Pavement Distress Data Collection*. Transportation Research Record: Journal of the Transportation Research Board, 1643: 80–85.

Transportation Research Board and National Cooperative Highway Research Program (2009). *NCHRP Synthesis 401 – Quality Management of Pavement Condition Data Collection, A Synthesis of Highway Practice.*

Transportation Research Board and National Cooperative Highway Research Program (2004). *NCHRP Synthesis 334 – Automated Pavement Distress Collection Techniques, A Synthesis of Highway Practice*.

University of New Mexico. (2009). 2009 Pavement Evaluation Report: Northern New Mexico. Albuquerque, NM.

US Department of Transportation: The Federal Highway Administration. (2003). *Distress Identification Manual for the Long Term Pavement Performance Program*. Washington DC.